## applied biosystems

# Chromosome Analysis Suite (ChAS) v4.3 USER GUIDE

Publication Number 702943 Revision 17





The information in this guide is subject to change without notice.

#### DISCLAIMER

TO THE EXTENT ALLOWED BY LAW, LIFE TECHNOLOGIES AND/OR ITS AFFILIATE(S) WILL NOT BE LIABLE FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE, OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING YOUR USE OF IT.

#### Important Software Licensing Information

Your installation and/or use of this Chromosome Analysis Suite software is subject to the terms and conditions contained in the End User License Agreement (EULA) which is incorporated within the Chromosome Analysis Suite software, and you will be bound by the EULA terms and conditions if you install and/or use the software.

#### NOTICE TO PURCHASER: DISCLAIMER OF LICENSE

Purchase of this software product alone does not imply any license under any process, instrument or other apparatus, system, composition, reagent or kit rights under patent claims owned or otherwise controlled by Life Technologies Corporation, either expressly, or by estoppel.

#### Legal entity

Affymetrix, Inc. | Santa Clara, CA 95051 USA | Toll Free in USA 1 800 955 6288

#### **TRADEMARKS**

All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.

©2021 Thermo Fisher Scientific Inc. All rights reserved.

## **Contents**

CHAPTER 1 Introduction	20
Features in v4.3	21
About this user guide	21
Customer support	21
CHAPTER 2 Installing ChAS	22
Recommended and minimum requirements	23
Requirements and prerequisites for arrays	24
Zip file contents	24
Installing ChAS	25
New Installation	25
Upgrade installation	25
Copying analysis files	26
Analysis file locations in Windows 10	26
Viewing Hidden Files and Folders	26
Displaying hidden files and folders in Windows 10	26
Analysis file download	26
Downloading ChAS analysis files from NetAffx for use with the ChAS Browser	27
Downloading ChAS analysis files from NetAffx using the Analysis Workflow	
Updating NetAffx Genomic Annotation files	28
Setting up a proxy server	28
Proxy server access to NetAffx from the analysis workflow	28
Proxy server access to NetAffx from the ChAS browser	29
Proxy Server access to a remote ChAS database server from the ChAS browser $\ldots$	29
Uninstalling	29
CHAPTER 3 Getting started	30
Starting ChAS	30
Logging into the ChAS database	31
Analysis workflow module	32
First time setup	32
Assigning an Input sample path(s)	33
Assigning an Output results path	34
Assigning a Central QC history path	
File types and data organization in ChAS	
ChAS file types	
File Types Supported in ChAS	
Region information files	

	Analysis and visualization library files	36
	Data organization in ChAS	
	Basic workflow for cytogenetics analysis	36
	Array processing workflow (using instrument control software)	
	Probe-level Analysis of CEL file data	
	Loading data into ChAS for display	39
	CytoScan array (CYCHP files)	39
	CytoScan XON array (XNCHP files)	
	CytoScan HTCMA array (RHCHP files)	
	Genome-wide SNP array 6.0 (CNCHP files)	
	OncoScan array (OSCHP files)	
	ReproSeq aneuploidy (zip files)	
	Viewing data and features of interest using the ChAS display controls	
	Working with ChAS	
	Accessing functions in ChAS	
	Changing pane sizes	
	Opening panes in separate windows	
CHA	PTER 4 CN/LOH/Mosaicism analysis	44
	Single sample analysis	
	ChAS analysis file compatibility	
	Introduction to single sample analysis	46
	Copy number segments on the X and Y chromosomes	
	Mosaic copy number segments on the X chromosome	
	LOH segments on X and Y chromosomes	
	CytoScan arrays	
	Performing a single sample analysis	
	Setting up and running a single sample analysis	
	Recentering CytoScan HD, 750K and Optima arrays	60
	Method 1	
	Method 2	
	No gender single sample analysis.	
	Setting up and running a normal diploid analysis	
	Setting up and running an OncoScan single sample analysis	
	Creating your own custom QC settings	
	Recentering OncoScan CNV and OncoScan CNV Plus arrays	
	Method 1	
	Method 2	
	Adding CEL files to analyze	
	Manually adding CEL files to analyze	
	CEL nie displaying options (optional)	

Importing CEL files using batch import	
Converting CEL files to CHP files	
How it works	
Supported array types	
Launching the tool	
Setting up the tool	
Assigning your input folder(s)	
Assigning your output folder(s)	
Assigning an archive folder	
Assigning a QC History file folder	
Selecting a target genome version	
Normal diploid analysis check box	
Running the tool	
Opening the newly generated CHP file(s)	
Reference files	
Creating a reference file	90
CHARTER E Analysis worldlow experts and OC tools	02
CHAPTER 5 Analysis workflow exports and QC tools	73
Displaying and exporting data from the analysis workflow	
Adding files to the QC results table	
Exporting QC table information	
Viewing analysis files in the ChAS browser	
Exporting probe-level data	
Exporting CytoScan Probe-Level Data	
Exporting OncoScan probe-level data	
Exporting a gene report (CytoScan or OncoScan arrays)	
Exporting a XON region report (CytoScan XON only)	99
Exporting a copy number expression overlap report	100
Exporting genotype data using the analysis workflow	
Exporting options	104
Exporting a SNP List	
Exporting a specific chromosome	
Selecting an output path and filename	
Multiple file output options	
Saving and importing attributes	
Exporting to Integrative Genomics Viewer (IGV)	
Principle component analysis	
PCA plot generation	
Sample display options	
Additional PCA graph display options	

Concordance checks	109
Performing a concordance check	109
Concordance table filter and display options	
Filtering	
Changing the view and/or order of sample and reference columns	
Exporting the currently displayed table	
Mendelian error checking	
Running an error checking analysis	
Interpreting an error checking analysis	
Analysis workflow troubleshooting	
Log rollover	
Log collection	118
CHAPTER 6 Loading Data	119
Introduction to loading data	
Loading files	122
Using the search feature	124
Copy number segment smoothing and joining (optional)	128
Viewing what segments were smoothed/joined	128
Setting smoothing and joining parameters	129
Segment data tab options and descriptions	130
About smoothing	131
Copy Number State for Smoothed Segments	
About joining	132
XON segment merging	133
Turning off XON merging	133
Setting QC parameters in the ChAS browser	134
Viewing QC thresholds	134
QC thresholds tab options and descriptions	136
Adding a QC property	139
Editing an existing QC threshold	139
Histogram data	140
Loading histogram data	140
Changing the default histogram filters	140
Changing the default histogram colors	141
Adding filtered histogram data	143
Removing a histogram	
CHAPTER 7 Viewing Data	145
Displaying options of analysis results data	145
Overview of ChAS window components	146

	Files list	
	Data types list	
	Named settings	150
	Status bar	
	Display area	151
	Upper panes	152
	Selected chromosome view	152
	Lower panes	152
	Changing the NetAffx genomic annotation file version	153
CH	HAPTER 8 Displaying Data in Graphic Views	154
	Graphic views	
	Karyoview	
	Whole genome view	159
	Displaying the Whole Genome View for a sample(s)	160
	Changing graph types	161
	Changing colors of the data points or line data	162
	Setting a default WGV state display and colors	163
	Creating WGV states	163
	Applying a saved WGV state(s)	164
	Deleting a saved WGV state	164
	Using the WGV zoom feature	
	Adding a reference line to the WGV	166
	Adding a comparison file	
	Selecting a new comparison file	169
	Exporting a WGV	169
	Selected chromosome view	170
	Detail view	172
	Viewing CytoScan XON data in detail view	173
	Annotation color codes	
	Annotation OMIM color codes	175
	Changing an annotation color	175
	Data in the detail view	
	Navigation controls in detail view	
	Obtaining summary metrics for a zoomed in region	
	Selecting a chromosome section for display	
	Karyoview and selected chromosome view	
	Selecting a chromosome for detailed examination	
	Examining a section of the chromosome	
	Coordinate range box	
	Going to a specific coordinate or coordinate range	18/

Returning to a previous location	
Zooming to a selected item	
Setting a vertical highlight	
Edit configurations in the misc tab	
Accessing the Misc tab	
Autosave	188
Coordinated box format	
Chromosome sorting order	
Zoom buffer	
CHP file colors	
Remapped segment patterns	190
Microarray nomenclature configuration	190
Controlling the display of data	
Selecting data for display	192
Selecting and deselecting files for display	
Changing the order of the Sample lanes or reference annotations	
Viewing data properties	
Closing a file	
Selecting data types for display	
Changing the grouping of samples and data types	
Changing the grouping	
Lanes grouped by sample	
Selecting a dark scheme display	
Expanding and contracting annotations	
Toggling between collapsed and expanded display of annotations:	
Changing graph appearance	
Opening the Graph Settings window	
Changing the settings for a graph type	
Selecting different graph styles	
Changing the graph style	200
Changing graph attributes	
Changing scale	
Changing the vertical height of a graph	
Learning more about features	
Pop-ups	
Turning pop-ups on or off	
Right-click menu options	
Selection details table	
Linking to external websites	
Linking to TaqMan copy number and genotyping assays	
Viewing and ordering TaqMan assays for CN	
Viewing and Ordering TagMan assays for genotyping	

CHAPTER 9 Filtering segments	. 219
Applying segment parameter filters	220
Opening the Segment Filters window	220
Using segment parameter filters	223
CHAPTER 10 Segment modification	. 225
Editing segment data overview	225
Using edit mode	226
Types of segment editing	227
Tracking original calls	228
Merging all segments types	228
Merging segment groups	228
Segment to segment merge	231
Un-Merging Merged Segments	
Deleting all segments types	
Un-deleting a deleted segment(s)	
Editing the start/end Coordinates of all segment types	
Un-doing the edited start/end coordinates of a segment	
De Novo segment drawing	
Un-doing a segment De Novo drawing	
Changing all copy number segment types	
Un-doing a copy number change	
Promoting mosaic segments	
Editing mode on	
Editing mode off	
Removing all edits made to a sample	
Editing the Microarray Nomenclature (ISCN 2013) and Microarray Nomenclature fields	247
CHAPTER 11 Sample and segment annotations	. 249
Sample annotations	249
Adding, removing, and changing the order of sample type text	249
Adding a Sample Type	250
Deleting a Sample Type	250
Re-arranging the order currently displayed Sample Types	250
Restoring the factory default Sample Types	
Adding, removing, and changing the order of phenotype text	250
Adding a short Phenotype text	251
Deleting a short Phenotype text	
Re-arranging the currently displayed Phenotypes order	
Restoring the factory default Phenotypes	251

Adding annotations at the sample (xxCHP) file level	251
Segment annotations	253
Setting up the calls feature	253
Adding and removing calls	253
Adding, deleting, and re-arranging calls	254
Adding calls to the Call drop-down list	254
Deleting calls from the Call drop-down list	254
Re-arranging the order of Calls in the Call drop-down list	254
Restoring the factory default Calls	
Using the calls feature	
Method 1: At the segments table	
Method 2: At the View/Edit Annotation Properties Window	
Adding or removing interpretation snippets	
Removing a saved snippet	
Removing multiple saved snippets	
Using the interpretation snippets feature	
Method 1: At the segments table	
Method 2: At the View/Edit annotation properties window	
Adding or removing inheritance calls	
Adding an Inheritance call to the Inheritance drop-down list	
Deleting a call from the Inheritance drop-down list	
Using the inheritance feature	
Method 1	
Method 2: From the View/Edit annotation properties window	
Adding Oncomine Reporter annotations	
Tracking and reviewing the log file	
CHAPTER 12 Using CytoRegions	268
CytoRegions overview	268
Selecting a CytoRegions information file	269
Viewing CytoRegions	271
CytoRegions in the graphic views	271
CytoRegions table	273
Highlighting regions in the CytoRegions table and details view	
CytoRegions tool bar	
Searching CytoRegions	
Using filters with CytoRegions	276
Using restricted mode	276
Assigning a CytoRegion for targeted XON analysis	
Creating an AED File from a gene list	
, and the second se	

CHAPTER 13 Using the overlap map and filter	280
Overlap map and filter overview	280
Using the overlap filter	280
Selecting the overlap map file	281
Viewing overlap regions	283
Viewing overlap map regions in the graphic displays	
Viewing the overlap map table	284
Highlighting overlap regions in the overlap map table and details view	284
Overlap map tool bar	
Using the overlap filter	
CHAPTER 14 Creating and editing AED files	288
Creating an AED file of annotations	288
Adding annotations to an AED file	290
Creating an AED file of regions	291
Creating a new CytoRegions file in AED file format	291
Adding regions to an existing AED file	293
Adding a section to a new region (AED) file	293
Deleting regions from an AED file	295
Viewing or Editing AED File Properties	295
Viewing file properties	295
Viewing the genome assembly version	296
Adding a property	297
Removing a property	298
Editing a property value	298
Viewing and editing annotations	298
Entering general information	300
Editing the General tab properties in AED annotations	300
Customizing properties	
Adding Properties	
Adding custom properties	
Adding a curation (Optional)	
New user annotations	
Viewing and batch editing AED file contents	
Protecting an AED file	
AED/BED color rules	
Selecting a new default color for loaded AED or BED files	
Creating a color rule	
Changing an annotation color	
Exporting information in AED or BED format	
Exporting position data as an AED or BED file	318

Merging and exporting feature position information for multiple files	320
Expression analysis AED file generation	321
Viewing a Gene Expression AED file in ChAS	321
CHAPTER 15 VCF files	322
Loading VCF files	322
Exporting VCF files	325
Exporting CN and/or variant data as a VCF file (for use in 3rd party browsers)	325
CHAPTER 16 Displaying data in table views	327
Display overview	327
Common table operations	327
Standard tool bar controls	328
Sorting by columns	328
Sorting a table by a single column	328
Performing a multi-column sort	
Changing the order of table columns	
Selecting columns to display or hide	
Using the right-click menu	
Using a column header's right-click menu	
Sum, mean, and median calculator	
Calculating multiple numeric values in a column	
Saved table states	
Saving the current segment table state to its default	
Adding columns to table states	
Removing (hiding) columns in a table (for report use)	
Saving a newly edited segment table state	
Applying previously saved table states	335
Deleting previously saved table states	
Segments table	
Segments table tool bar	
Segments table	
Segment table columns	
Obtaining all annotations associated with a segment	
Graphs table	
Graphs table properties	
Graph Table Tool bar	
To export genotype calls	
Variants table	
Tool bar	
Variants table	352

CytoScan HTCMA	352
OncoScan CNV Plus	
QC and sample info tab	
QC and sample information table	
CytoScan QC view	
Default QC view	
CytoScan HTCMA QC view	
OncoScan QC view	
ReproSeq QC view	
SMN Sample Info view	
<u> </u>	
Loaded AED/BED files table	
Chromosome summary data	
Auto-generated Autosome LOH percentage	
Searching results	
Searching within a selected file	
Finding intersections	369
CHAPTER 17 Prioritizing segments	373
Tier-based prioritization	373
Configuring the tier-based option	373
Tier to call settings	
Viewing segment prioritization in the segments table	
Score-based prioritization	
Configuring the score-based option	
View/Edit Score-Based Rules for Gain Segments	
View/Edit Score-Based Rules for Loss Segments	
Configuring the score-based option	
Viewing segment prioritization in the segments table	
CHAPTER 18 Interacting with the ChAS database	387
Setting up a ChAS DB guery	388
Setting up query parameters for a copy number search	
Setting up query parameters for an LOH segment search	
Setting up query parameters for an XON region segment search	
Querying a segment from the segment table	
Segment intersections	
-	
Additional segment intersection information	
Downloading segments from a sample file in ChAS DB	
Updating downloaded segment annotations	399

Filtering DB count columns	
Querying a segment from the detail view	401
Changing or refining the DB query criteria	
Publishing data to the database	404
Publishing data or multiple data to the database	404
Method 1	404
Method 2	
Publishing to database	
Important rules and restrictions	
Manual or automatic connection mode	
Querying samples in the ChAS database	
Editing column contents	
Editing multiple cells with the same value	
Removing a sample from the query window	409
Deleting sample(s) from the ChAS database	409
Deleting a single sample	409
Deleting multiple samples	
Querying segments to the ChAS database	
Editing column contents	
Editing multiple cells with the same value	412
Exporting graphic views	
Exporting as a PDF	
Creating signature and background profiles	
Signature profiles	
g p	
Creating a new signature	
Creating a new signature	
Creating a new signature  Editing a saved Signature  Deleting a signature	416
Editing a saved Signature	
Editing a saved Signature  Deleting a signature	
Editing a saved Signature  Deleting a signature  Background profiles	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background  Editing a saved background  Deleting a background  Exporting as Word (DOCX) format	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background  Editing a saved background  Deleting a background	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background  Editing a saved background  Deleting a background  Exporting as Word (DOCX) format	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background  Editing a saved background  Deleting a background  Exporting as Word (DOCX) format  Exporting as PNG	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background  Editing a saved background  Deleting a background  Exporting as Word (DOCX) format  Exporting as PNG  Exporting table data	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background  Editing a saved background  Deleting a background  Exporting as Word (DOCX) format  Exporting as PNG  Exporting table data  Exporting table data into a PDF  Exporting tables tips	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background  Editing a saved background  Deleting a background  Exporting as Word (DOCX) format  Exporting as PNG  Exporting table data  Exporting table data into a PDF  Exporting table data into a PDF file	
Editing a saved Signature  Deleting a signature  Background profiles  Creating a new background  Editing a saved background  Deleting a background  Exporting as Word (DOCX) format  Exporting as PNG  Exporting table data  Exporting table data into a PDF  Exporting tables tips	

Exporting a segments table with modified segments to a TXT file	425
Transfer to clipboard	426
Combining PDFs into a single PDF	428
Adding a new PDF export to an existing PDF file	428
Combining two existing PDF files	429
Exporting with ClinVar	
Creating a ClinVar submission profile	
Editing an existing ClinVar profile	433
Deleting a profile	
Adding vocabulary content	
Removing vocabulary content	
Exporting in ClinVar format	
Exporting a deletion log	
CHAPTER 20 User profiles and named settings	436
Types of settings	436
User profiles	436
Named settings	
Creating and using user profiles	
Deleting a user profile	
Creating and using named settings	
Selecting a named setting	
Exporting and importing preferences	
Exporting preferences	
Importing preferences	
Importing hyperlinks	444
Configuring the HTTP service	
CHAPTER 21 Database tools	445
Connecting to a remote ChAS DB server	445
Accessing the ChAS DB server tools	
Status page	
Maintenance	
Backing up a database	449
Restoring a database	450
Merging ChAS databases	451
Merging two ChAS databases	451
Merging an older ChAS database	
Cleaning up a database	452
Downloading deletion logs	452

Creating a blank ChAS DB	452
Deidentifying Files	453
Administration	454
Permission Guidelines	455
Using a shared ChAS database while off-line	455
Working off-line using a ChAS database that resides on a shared server	455
Publishing data you have analyzed in off-line mode to the shared ChAS DB server	
Remapping a hg19 ChAS DB to hg38	456
CHAPTER 22 ChAS Database Loader (CDL)	<b>/.5</b> 9
CHAPTER 22 CHAS Database Loader (CDL)	. 450
Starting CDL	459
Adding files to CDL	460
Sample info	461
Adding files to be published using a text file	463
Publishing to the ChAS database	463
Testing your connection (optional)	463
Verifying the ChAS database	463
Before publishing files	464
Changing segment filters (optional)	464
Managing data types (optional)	465
Publishing your files	466
Clearing Table Data	469
Clear Published	469
Clear All	469
Closing CDL	469
APPENDIX A Analysis parameters	. 470
Analysis parameters for single sample analysis	7.70
Reference model file creation	
Neter effice model file creation	4/0
APPENDIX B AED file format	. 471
AED file description	471
Header row	472
Metadata records	473
Annotations Rows	474
Property name elements	474
Namespaces	
Properties	
AFFX properties	475
AFD properties	/,75

Biology properties	476
Style properties	
Types	
Compatibility	
UCSC Browser Extensible Data (BED)	478
Microsoft Excel and other spreadsheet applications	479
Microsoft Notepad and other Text editors	479
Text Editors	
References	479
APPENDIX C ChAS properties and types	481
Identifying properties within ChAS	481
Standard AED property style	481
ChAS property style	482
Full URI property style	482
Identifying value types within ChAS	483
Automatic conversion of xxCHP headers to properties	483
Header name conversion	483
Converted properties	484
Derived properties	484
APPENDIX D ChAS browser NetAffx Genomic Annotations	486
Homo Sapiens database files	486
NetAffx Genomic Annotation files	486
Source of content	486
APPENDIX E Genomic position coordinates	487
Genome assemblies	487
SNP and marker positions	487
Segment positions	488
APPENDIX F Editing BED files	489
APPENDIX G CytoScan algorithms and QC metrics	491
Algorithm overview	491
Feature identification and signal extraction	491
Single sample CytoScan workflow	491

Signal-level covariate adjustors	494
Fragment adapter covariate adjustor	494
Fragment length covariate adjustor	
Dual quantile normalization	
Copy number workflow	
Log2 ratio calculation	
High pass filter image correction	
Log2 ratio-level covariate adjustors	
Super GC covariate adjustor	
Reference intensity covariate adjustor	
Marker Type Covariate Adjustor	
Median Autosome normalization	
Systematic residual variability removal	
Signal restoration	
Segmentation	
Copy Number Calls for each Marker based on Log2 Ratios	
Segment Formation	
Smoothing & Joining	
Segment table output	
Mosaicism segment algorithm	
Limitations in mosaicism segmentations	
SNP marker workflow	
Signal summarization	499
Allelic signal computation	499
Genotyping	499
Allelic difference GC correction	499
Detection of LOH	
Array data QC metrics	500
Median of the Absolute values of all Pairwise Differences (MAPD)	
Waviness SD	
SNPQC	500
ndSNPQC (SNP Quality Control of Normal Diploid Markers)	
ndWavinessSD (Normal Diploid Waviness Standard Deviation)	
MAPD - Detailed Description	
Effect of MAPD on functional performance	
Waviness-SD – Detailed Description	
SNPQC - Detailed Description	
Effect of SNPQC on Functional Performance	
TuScan algorithm	
Manual re-centering algorithm (OncoScan)	
Copy number effect on somatic mutations	
CytoScan XON region calling algorithm	

APPENDIX H Recommended CytoScan XON array workflows		
Whole exome analysis	510	
Recommended workflow for analyzing the whole exome	510	
Targeted analysis	512	

## 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 CytoScan<sup>™</sup> Arrays and OncoScan Arrays.
- Analyze segment data at different levels of resolution.
- View results data (CYCHP, XNCHP, and OSCHP files) that summarize chromosomal aberrations in table and graphical formats.
- Display CNCHP data from Genome-Wide Human SNP 6.0 Array.
- Customize and load your own annotations and regions for focused analysis.
- Display ReproSeq Aneuploidy data from Ion Reporter<sup>™</sup>.
- 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 UCSC Genome Browser, NCBI, DECIPHER, ClinVar, and ClinGen.
- Export user-selected data in graphical and tabular formats.
- Store and query segment data for streamlined analysis.
- Check for Mendelian Inheritance errors across related samples.
- Combine gene expression and copy number data in both tabular and graphical formats.
- Export data for viewing in Integrative Genome Viewer (IGV).
- Directly access the TaqMan website for follow up analysis.

**IMPORTANT!** The results from ChAS are for Research Use Only and not for use in diagnostic procedures.

Chromosome Analysis Suite is not a secondary analysis package. However, it does create CYCHP, OSCHP, XNCHP, and tab-delimited text files required for secondary analysis packages.

#### Features in v4.3

- Improved mosaic calling algorithm.
- Updated VCF visualization to include copy number segmentation.
- Enable automatic CEL analysis from multiple input folders.
- Added OMIM Region Phenotype annotations which completes all annotations in the OMIM Morbidity Map.

#### About this user guide

This user guide provides step-by-step instructions for performing the procedures required to use ChAS and can also be accessed from the software by clicking on the top menu bar's **Help** drop-down.

The steps outlining procedures through out this User Guide are frequently supplemented with screen captures to further illustrate the instructions.

**Note:** The screens that were captured for this User Guide may not exactly match the windows displayed on your screen.

#### **Customer support**

Visit thermofisher.com/support for the latest in service and support, including:

- Worldwide contact telephone numbers
- Product support, including:
  - Product FAQs
  - Software, patches, and updates
- Order and web support
- Product documentation, including:
  - User guides, manuals, and protocols



## Installing ChAS

#### This chapter includes:

- "Recommended and minimum requirements" on page 23
- "Requirements and prerequisites for arrays" on page 24
- "Zip file contents" on page 24
- "Installing ChAS" on page 25
- "Viewing Hidden Files and Folders" on page 26
- "Setting up a proxy server" on page 28
- "Uninstalling" on page 29

ChAS is a stand-alone application that supports the analysis and/or visualization of the following results data files:

- CytoScan (CYCHP)
- OncoScan (OSCHP)
- CytoScan XON (XNCHP)
- Genome-Wide Human SNP 6.0 (CNCHP)
- ReproSeq Aneuploidy (.zip)
- BED/AED files
- VCF files

**IMPORTANT!** Due to the amount of memory that ChAS requires to operate, Thermo Fisher Scientific VERY STRONGLY recommends that you DO NOT install the ChAS software on instrumentation computers being used for scanning and operating fluidics systems.

## Recommended and minimum requirements

**Note:** The full Database and Browser software must be installed on at least one or more 64-bit analysis workstations to create results data files.

Table 1 Software

System Properties	Recommended System Requirements	Minimum System Requirements		
Processor	3 GHz (or greater) Pentium Quad Core Processor	3 GHz (or greater) Pentium Dual Core Processor		
64-bit Windows <sup>®</sup> Operating System and Web Browser	Windows 10 Internet Explorer v.11 or above	Windows 10 Internet Explorer v.11 or above		
Available Disk Space	250 GB HD + data storage	150 GB HD + data storage		
Free Disk Space Required at Installation	≥ 5 GB	≥ 5 GB		
RAM	16 GB	8 GB		

Table 2 Server

System Properties	Recommended System Requirements	Minimum System Requirements		
Processor	3.1 or 3.3 GHz Quad core processor	2.7GHz Quad core processor		
Windows Operating System	Windows 2012 Server R2 Standard 64-bit	Windows 2012 Server R2 Standard 64-bit		
Available Disk Space	1 TB HD + data storage	512 GB HD + data storage		
Free Disk Space Required at Installation	<u>&gt;</u> 5 GB	<u>&gt;</u> 5 GB		
RAM	≥24 GB	16 GB		

#### Requirements and prerequisites for arrays

IMPORTANT! A Windows 64-bit Operating System is required for all array types.

**IMPORTANT!** Chromosome Analysis Suite requires AGCC 4.3/GCC 6.1 or higher to produce CytoScan/Oncoscan CEL files.

**IMPORTANT!** 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 ~155 MB. A CytoScan XON Array XNCHP files is ~174MB.

**IMPORTANT!** The ChAS software has been verified for use on a Windows 64-bit Operating System. ChAS may work on other Windows Operating Systems, but only the 64-bit version has been verified.

#### Zip file contents

Go to thermofisher.com ChAS product page to downloaded the zip files.

To analyze the CytoScan HD array CEL files and visualize the data in ChAS, you must download the following ChAS 4.3 files:

- Chromosome\_Analysis\_Suite\_4.3.zip
- CytoScanHD\_Array\_Analysis\_Files\_NA33.r9.zip

Before performing an analysis, you must download the appropriate zip file(s) listed below:

- CytoScan 750K: CytoScan 750K\_Array\_Analysis\_Files\_NA33.r9.zip (and/or) CytoScan 750K\_Array\_Analysis\_Files\_NA36.r6.zip
- CytoScan Optima: CytoScanOptima\_Array\_Analysis\_Files\_NA33.r9.zip (and/or) CytoScanOptima\_Array\_Analysis\_Files\_NA36.r6.zip
- OncoScan CNV: OncoScan\_CNV\_Analysis\_Files\_NA33.r2.zip and/or NA36.r2.zip
- OncoScan CNV Plus Assay: OncoScan\_Array\_Analysis\_Files\_NA33.r6.zip and/or NA36.r2.zip
- CytoScan XON: CytoScanXON\_Array\_Analysis\_Files\_hg19.r4.zip and/or hg38.r4.zip

After downloading and extracting the **Chromosome\_Analysis\_Suite\_4.3zip** file, a **Chromosome Analysis Suite 4.3** folder appears containing the following files:

- ChAS4.3\_setup.exe
- ChAS\_4.3\_Manual.pdf
- ChAS\_4.3\_Release\_Notes.pdf
- aDGV (HD + XON) ChAS DB hg19 backup.db files
- aDGV (HD+XON) ChAS DB hg38 backup.db files

#### Installing ChAS

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

#### **New Installation**

1. Double click on the **ChAS4.3\_setup.exe** file from the "Chromosome Analysis Suite 4.3" folder.

The Install Shield Wizard for Chromosome Analysis Suite begins.

2. At the Welcome window, click Next.

The License Agreement window appears.

3. Please read the license agreement carefully, click the "I accept the terms of the license agreement" radio button, then click **Next**.

The Setup Type window appears.

4. Click the appropriate drive's check box, then click Next.

The PostgreSQL Database Server Installation window appears.

**Note:** Make sure you select a drive with the most available space. Consider 1 GB of space is required for every 1000 samples you add to the database.

The installer auto-detects and displays a default **Port** number. It is recommended that you do not change this number.

5. Click Next.

The Start Copying Files window appears.

- 6. Click **Next**, then follow the on-screen instructions to complete the installation.
- 7. After all software installation is complete, you must download the new Analysis files from the NetAffx site or copy them into your ChAS Library folder. If you are unable to connect to the Internet, refer to "Copying analysis files" on page 26.

**IMPORTANT!** If your Windows Firewall is enabled during the installation of ChAS and you want to Backup the ChAS Database and Restore it to your local ChAS DB (see "Using a shared ChAS database while off-line" on page 455) a message may appear indicating that you cannot connect to the shared folder.

If this message appears, contact your IT department for help in allowing file sharing through the Windows Firewall.

## Upgrade installation

**IMPORTANT!** The ChAS 4.3 Installer does NOT support upgrade from previous versions. Due to an updated version of the ChAS DB, previous versions of ChAS must be uninstalled using add/remove programs prior to running the ChAS 4.3 installer. Be sure to make a backup of your ChAS DB prior to uninstalling your previous version of ChAS.

To keep current preferences, see "Exporting and importing preferences" on page 442.

## Copying analysis files

The CytoScan Analysis Library Files zip package download contains the Analysis Files required to process their respective CytoScan Array CEL files into CYCHP files.

The OncoScan Analysis Library Files.zip package download contains the Analysis Files required to process their respective OncoScan Array CEL files into OSCHP files.

The CytoScan XON Analysis Library Files.zip package download contains the Analysis Files required to process their respective CytoScan XON Array CEL files into XNCHP files.

Also included in the CytoScan HD Analysis Library Files. zip are the files for GenomewideSNP\_6 files (required in ChAS to view GenomeWideSNP\_6 CNCHP files).

If you are unable to download library files through the software, you can download the zipped library files from the ChAS product page at www.thermofisher.com, then extract (unzip) the files into the following location: C:\Affymetrix\ChAS\Library

# Analysis file locations in Windows 10

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

#### Viewing Hidden Files and Folders

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

# Displaying hidden files and folders in Windows 10

- 1. At the Windows 10 Desktop, move your mouse to the bottom right of the Task bar (right of the clock).
  - Five large icons appear.
- 2. Click on the **Settings** icon.
- 3. Click Control Panel.
  - The Control Panel window opens.
- 4. Click **Appearance and Personalization** in Control Panel. Under Folder Option, click "Show hidden files and folders".
- 5. In the Folder Options window 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.
- Click **OK**.
- 7. Close all open windows.

## 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" on page 438)

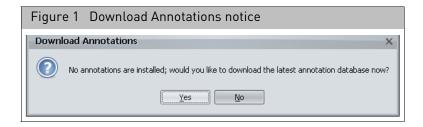
**Note:** To process the CytoScan Arrays in AGCC/GCC, you must install the appropriate library files for AGCC/GCC on the AGCC/GCC workstation (see the specific array product page at www.thermofisher.com for details).

You can download the ChAS analysis files from NetAffx using either the ChAS Browser or the Analysis Workflow. The files will be saved into the same folder whether downloading through the ChAS Browser or the Analysis Workflow.

#### Downloading ChAS analysis files from NetAffx for use with the ChAS Browser

- 1. Create a NetAffx account with a user name and password. To do this, go to thermofisher.com, click **Sign In**, then click **Register** to create an account.
- Start ChAS.

If no annotations are installed, a Download Annotations notice appears. (Figure 1)



#### 3. Click Yes.

The NetAffx Download Service window opens.

**Note:** You can also open the NetAffx Download Service window by selecting **Update Library and Annotation Files** from the Help menu.

- Enter your NetAffx user name and password, then click Next.
   The Select the Library Files to Download window opens with a list of library files that
  - your local hardware can support.
- 5. Select the library and annotation files you want to download.
- 6. Click **Download**.

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

- 7. Click **OK** when the download is complete.
- 8. The NetAffx Authentication window remains open, click **Cancel** when finished downloading the library files.

#### Downloading ChAS analysis files from NetAffx using the Analysis Workflow

- 1. Create a NetAffx account with a user name and password. To do this, go to thermofisher.com, click **Sign In**, then click **Register** to create an account.
- 2. Select Analysis → Perform Analysis Setup.
- 3. Select Utility Actions  $\rightarrow$  Download library Files.
- 4. Enter your NetAffx user name and password, then click **Next**.
- 5. Select the array type check box(es) for the analysis files that you would like to download.
- 6. Click Download.

#### Updating NetAffx Genomic Annotation files

The publicly available annotations in the NetAffx Genomic Annotation file are updated quarterly. When launching the ChAS Browser or Analysis Workflow, the software will check for the availability of new NetAffx Genomic Annotation files.

If more current files are available, you will be prompted to download the files from NetAffx using the instructions from either of the previous two sections.

If you are unable to download from NetAffx through the software, please contact Technical Support for access to the current NetAffx Genomic Annotation file.

#### Setting up a proxy server

**Note:** If you do not know what the proxy settings are, contact your IT department.

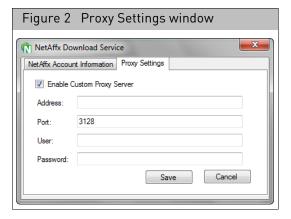
Follow the steps below if your system has to pass through a Proxy Server before it can access the NetAffx server.

To change from a Local Disk (C: or D:) Database to a dedicated remote Database Server, see "Proxy Server access to a remote ChAS database server from the ChAS browser" on page 29.

Proxy server access to NetAffx from the analysis workflow

- 1. Launch the Analysis Workflow by selecting **Analysis** → **Perform Analysis Set** Up in the ChAS Browser.
- Click Utility Actions → Download Library Files.
   The NetAffx Download Service window appears.
- 3. Click the **Proxy Settings** tab. (Figure 2)
- 4. Click the **Enable Custom Proxy Server** check box.
- 5. Enter the Host Address, Port (if not listed), User and Password.

**IMPORTANT!** This proxy user ID and password is NOT the same ID and password used to connect to NetAffx.



6. Click Save.

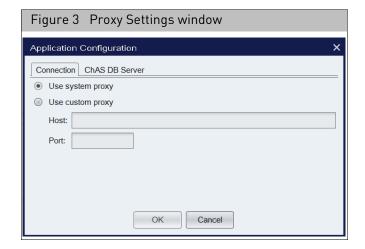
# Proxy server access to NetAffx from the ChAS browser

- 1. From the **Help** drop-down menu, click **Update Library and Annotation Files...**The NetAffx Download Service window appears.
- 2. Click the **Proxy Settings** tab. (Figure 2)
- 3. Click the **Enable Custom Proxy Server** check box.
- 4. Enter the Host Address and Port information, then enter your user name and password.

**IMPORTANT!** This proxy user ID and password is NOT the same ID and password used to connect to NetAffx.

5. Click Save.

Proxy Server access to a remote ChAS database server from the ChAS browser 1. From the **Preferences** drop-down menu, click **Edit Application Configuration...**The Configuration window appears. (Figure 3)



- 2. Click the **Use custom proxy** button.
- 3. Contact your IT department for help with entering the Host and Port information.
- 4. After you have completed the appropriate fields, click **OK**.

#### Uninstalling

**IMPORTANT!** It is strongly recommended you backup your database BEFORE uninstalling ChAS.

- 1. From the Windows Start Menu, navigate to the Windows Control Panel,
- 2. Select Uninstall or change a program.
- 3. Locate the **Chromosome Analysis Suite** application, then perform the uninstall as you normally would.
- 4. Click **OK** to acknowledge the message box that warns the ChAS application must be closed (before removing it).



## Getting started

#### This chapter includes:

- "Starting ChAS"
- "Logging into the ChAS database"
- "File types and data organization in ChAS" on page 34
- "Basic workflow for cytogenetics analysis" on page 36
- "Working with ChAS" on page 42

#### Starting ChAS

1. Double-click on the Desktop ChAS icon.
The **Select User** window appears. (Figure 4)



2. Use the drop-down button to select a user or click **Create New** to create a new user profile. For more information see "Creating and using user profiles" on page 438.

ChAS

- 3. Optional: Click the Manual connection check box. For information on manual connections, see "Manual or automatic connection mode" on page 406.
- 4. Click OK

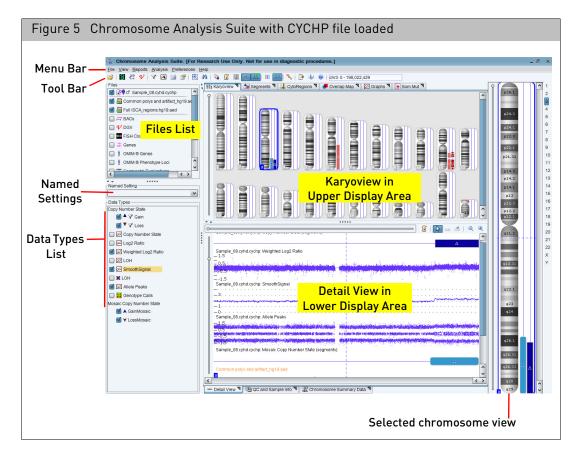
The Chromosome Analysis Suite application opens. as shown in Figure 5 on page 31 after logging into the ChAS DB. To login, see "Logging into the ChAS database" (below).

**Note:** A message may appear indicating a more current version of the NetAffx Genomic Annotation file is available for download. To download the newer version of the file, see "Analysis file download" on page 26. If you are unable to download the files via the NetAffx dialog, please contact Technical Support for alternative downloading options.

#### Logging into the ChAS database

See Chapter 21, "Database tools" on page 445 for steps on how to log into the ChAS Database.

**Note:** The ChAS installation automatically generates a default user name (admin) and password (admin). To change these default passwords and/or create additional user names, see "Administration" on page 454. Also, your sign in (login) information is retained throughout a working session, however, if the ChAS application is closed, then opened again, you must login again.



The ChAS browser window has the following components:

- Menu Bar Provides access to the functions of the software.
- Tool Bar Provides quick access to commonly used functions. Note: Some features that were previously in the Tool bar (such as Dark/Light Schema) have been removed, but continue to be available under the View menu item.
- **Files List** Shows the data and annotation files that are available for display. See "Files list" on page 147.
- **Data Types List** Displays the type of data available in the files. See "Data types list" on page 149.
- Named Settings Displays a list of the previously saved display settings for ChAS. See "Named settings" on page 150.

- Status Bar 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. See "Status bar" on page 150.
- Display Area Displays the following data in graphical and table formats:
  - CYCHP, CNCHP, XNCHP, OSCHP, RHCHP, and/or ReproSeq Aneuploidy data
  - Detected segments
  - Region information file data
  - Histogram data (representing segments uploaded to the database)
  - Reference annotations

For more display area information, see "Viewing Data" on page 145.

#### Analysis workflow module

The Analysis Workflow generates xxCHP files from CEL files and tracks ongoing ChAS analysis tasks.

You can access the Analysis Workflow at any time by clicking on its @ ic

#### First time setup

After installation, you must configure your data paths.

The Analysis Workflow requires the following steps:

1. From the Analysis menu, select **Perform Analysis Setup.** (Figure 6)



The Analysis Workflow Configuration window tab appears. (Figure 7)

Save Cancel

# ### Please select output results path Pl

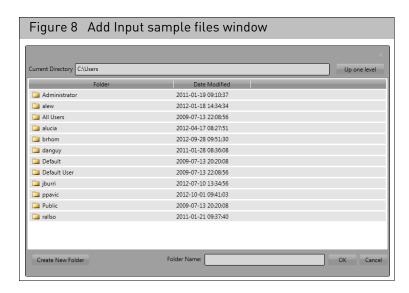
If it does not appear, click **Utility Actions**  $\rightarrow$  **Configuration**.

## Assigning an Input sample path(s)

A minimum of two sample paths is recommended.

1. Click Add.

The Add Input sample files window appears. (Figure 8)



Click the Up one level button to navigate to the recommended C:\Users directory, then click the Create new folder button to label a new input sample folder.

(Example: C:\Users\YourUserName\CytoScan Data)

3. Repeat **Step 2** to create a second input folder in an easy to access area. (Example: C:\Users\YourUserName\Collaborator\_OncoScan Data)

## Assigning an Output results path

1. Click the **Browse** button.

A good practice is to navigate to your current ChAS data folder location, then click **Create New Folder** to create an output results path folder <u>inside</u> this folder. (Example: C:\Users\YourUserName\CytoScan Data\CytoScan\_Results\_Files)

2. Click Save.

## Assigning a Central QC history path

- 1. Click the **Browse** button, then navigate to a folder in which to store the QC history file. (Example: C:\Cytoscan\_data\)
- 2. Click **Create New Folder** to create a central QC history path folder. (Example: QC\_History)

#### File types and data organization in ChAS

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

#### ChAS file types

ChAS uses the following types of files:

- Data files
- Region Information files
- Support files

#### File Types Supported in ChAS

Some data files that ChAS uses are generated by other Thermo Fisher Scientific software, as shown in Table 3.

Table 3 Supported file types

File Type	Created In	ChAS	
Sample file (ARR)	AGCC/GCC	Uses this information to associate sample attribute information with CEL and xxCHP or CNCHP files.	
Intensity Data file (CEL)	AGCC/GCC	Analyzes the intensity data in the CEL file, then generates a xxCHP files.  Note: A 64-bit system is required to analyze intensity data.	
Analysis Results (CYCHP)	ChAS	Displays results in graphical and tabular formats.	
CytoScan array: CYCHP contains copy number, LOH, mosaicism, and genotype call information			
Analysis Results (RHCHP from HTCMA) contains copy number, LOH, Carrier Variant, and SMN results	RHAS	Displays results in graphical and tabular formats.	
Analysis Results (OSCHP) OncoScan FFPE and OncoScan_CNV arrays, OSCHP contains copy number/LOH and somatic mutation information (FFPE Assay only)	ChAS	Displays the probe-level analysis, segment level data, and somatic mutation data.	

Table 3 Supported file types

File Type	Created In	ChAS
Analysis Results (XNCHP) CytoScan XON array: XNCHP contains LOH, Exon Region, and Genotype information.	ChAS	Displays probe-level analysis, segment level data in graphical and tabular formats.
Analysis Results (CNCHP) GenomeWide SNP 6.0 array contains copy number and LOH segments	GTC	Displays probe-level analysis data and generates segment data on-the-fly.
CHP Change Archive (CHPCAR)	ChAS	Stores user-annotated segment and sample annotations as well as modifications made to the segment data.
Analysis Results (.zip) ReproSeq Aneuploidy contains copy number and tiling information	ION Reporter	Displays copy number segment data and sequence tile information.
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, and DOCX)	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.
VCF Files	Other software packages	Allows users to view genotype and indel data in the Detail View.

#### 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**  $\rightarrow$  **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 use the reference annotations to provide region information or use the Export feature to export data in existing BED files to an AED file. See "Exporting information in AED or BED format" on page 318.

# Analysis and visualization library files

#### **IMPORTANT!** Every feature in ChAS requires support files.

- Array-type specific Library file sets with files for running Copy Number/LOH/ Mosaicism analysis and Reference Creation workflows (Analysis files)
- Files for visualizing and exporting data from xxCHP results data files.
- Reference Annotation files
  - Browser Annotation files are named using the following format:
     NetAffxGenomicAnnotations.Homo\_sapiens.hgXX.naYYYYMMDD.db>

## Data organization in ChAS

ChAS enables you to keep your CEL and Analysis Results files in any folder on your computer. As long as you know where the files are, you can load them from anywhere and move them around at your convenience.

**IMPORTANT!** It is recommended that you perform analysis operations with all analysis files stored on a local disk drive.

#### Basic workflow for cytogenetics analysis

**Note:** xxCHP is used when referring to CYCHP, CNCHP, XNCHP, and OSCHP files.

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

ChAS can be used to:

- Perform probe-level analysis of CEL file data for CytoScan and OncoScan Arrays.
- Display probe-level analysis data (xxCHP) from:
  - CytoScan arrays (CYCHP)
  - Genome-Wide Human SNP array 6.0 (CNCHP)
  - OncoScan arrays (OSCHP)
  - CytoScan XON arrays (XNCHP)
  - CytoScan HTCMA arrays (RHCHP)

**Note:** There are some differences in the way the ChAS handles these different types of arrays and how it treats the data from these four types of files. The basic cytogenetic analysis workflow includes the following steps:

- "Array processing workflow (using instrument control software)" on page 37.
- "Probe-level Analysis of CEL file data" on page 37.
  - For CytoScan arrays, this analysis is performed in CHAS and produces CYCHP or XNCHP files depending on array type. See "CN/LOH/Mosaicism analysis" on page 44.

- For CytoScan HTCMA arrays, analysis is performed in the RHAS and produces RHCHP files.
- For the Genome-Wide Human SNP Array 6.0, this analysis is performed in Genotyping Console (GTC) software and produces CNCHP files. For more details, please refer to the GTC User Guide.
- For the OncoScan files, this analysis can be performed in ChAS and produces OSCHP files.
- "Viewing data and features of interest using the ChAS display controls" on page 41.

# Array processing workflow (using instrument control software)

Array processing is performed in AGCC 4.1.2/GCC 5.0 or higher for the CytoScan Arrays, OncoScan Arrays, and Genome-Wide Human SNP 6.0 Array.

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

The array processing includes the following steps:

- 1. Registering samples and arrays.
- 2. Washing and staining the arrays.
- 3. Scanning arrays and generating intensity (CEL) file data.

The following file types are produced:

- Sample (ARR files)
- DAT Files
- CEL Files
- Audit
- JPG

See the Instrument Control Console Users Guide for more information.

#### Probe-level Analysis of CEL file data

Copy number 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 CEL files.

- For CytoScan arrays, this analysis is performed in ChAS and produces CYCHP or XNCHP files (depending on array type) and contain the data shown in Table 4. See "CN/LOH/Mosaicism analysis" on page 44.
- For CytoScan HTCMA arrays, the analysis is performed in the RHAS which produces a RHCHP file for viewing in ChAS.
- Genome-Wide Human SNP Array 6.0 Data: The probe level analysis on CEL file data is performed in GTC and produces the CNCHP file data types shown in Table 4. See the GTC User Guide for more information.

 For OncoScan Data: The probe level analysis on CEL file data is performed in ChAS and produces the OSCHP file data types show in Table 4.

Table 4

	Analysis Results *			
	CytoScan Array <sup>1</sup>	Genome-Wide Human SNP Array 6.0 <sup>2</sup>	OncoScan <sup>3</sup>	CytoScan XON <sup>4</sup>
Graph Data for the individual CN and SNP probes				
Copy Number State	yes	yes	yes	no
Log2 Ratio	yes	yes	yes	yes
Weighted Log2 Ratio	yes	no	yes	yes
LOH	yes	yes	yes	yes
Allele Difference	yes	yes	yes	yes
Genotype Calls	yes	no	no	yes
Smooth Signal	yes	yes	yes	yes
Variant Data	no	no	yes (CNV Plus only)	no
B-allele Frequency	yes	no	yes	yes
Segment data				
Gain and Loss segments based on runs of aberrant Copy Number State data	yes	yes	yes	no
Mosaic Gain and Loss segments of non-integer Copy Number States between CN=1 and CN=3	yes	no	no	no
LOH (Loss of Heterozygosity) based on runs of SNPs where heterozygote calls are absent	yes	yes	yes	yes
Exon Region Gain and Loss segments	no	no	no	yes

<sup>\*</sup> CYCHP for CytoScan, CNCHP for Genome-Wide Human SNP 6.0 Array, XNCHP for CytoScan XON Array, and OSCHP for OncoScan Array.

**Note:** Segment types drawn with flat ends (Gain and Loss) are the result of algorithms which can ascertain precise marker-to-marker breakpoints. Segment Types drawn with rounded ends (LOH, GainMosaic, LossMosaic) are the output of algorithms which closely approximate breakpoints based on the data.)

<sup>1)</sup> For more details on CytoScan Array data, see *Table 13 on page 177*.

<sup>2)</sup> For more details on Genome-Wide SNP Array 6.0 data, see *Table 16 on page 179*.

<sup>3)</sup> For more details on OncoScan FFPE Assay Data, see *Table 17 on page 180*.

<sup>4)</sup> For more details on CytoScan XON array data, see Table 15 on page 178.

# Loading data into ChAS for display

You perform the same steps for the different types of analysis data (CYCHP, CNCHP, XNCHP, OSCHP and .zip (ReproSeq Aneuploidy), but ChAS handles these types of data differently.

#### CytoScan array (CYCHP files)

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

- 1. Selects the 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)
    - Mosaic Gain/Loss
  - Graph Data
    - Copy Number State
    - Log2 Ratio
    - Weighted Log2 Ratio
    - · Smooth Signal
    - Loss of Heterozygosity (LOH)
    - Allele Difference
    - B-allele Frequency
    - Genotype calls

#### CytoScan XON array (XNCHP files)

When loading XNCHP files into ChAS for viewing the software:

- 1. Selects the segments in the XNCHP file to display as segments.
- 2. Displays the segments and graph data:
  - •Segment Data
    - Loss of Heterozygosity
    - XON Region Gain/Loss
  - Graph Data
    - Log2 Ratio
    - Weighted Log 2 Ratio
    - Smooth Signal
    - Loss of Heterozygosity (LOH)
    - Allele Difference
    - B-allele Frequency (BAF)
    - Genotype Calls

#### CytoScan HTCMA array (RHCHP files)

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

- 1. Selects the segments in the RHCHP file to display as segments.
- 2. Displays the segments and graph data:
  - Segment Data
    - Copy Number Gain/Loss
    - Loss of Heterozygosity (LOH)
  - Graph Data
    - Copy Number State
    - Log2 Ratio
    - Smooth Signal
    - Loss of Heterozygosity (LOH)
    - Allele Difference
    - B-allele Frequency
    - Variant Data

#### 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)

#### OncoScan array (OSCHP files)

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

- 1. Selects the segments in the OSCHP file to display as segments.
- 2. Displays the segments and graph data:
  - · Segment data
    - Copy Number Gain/Loss
    - Loss of Heterozygosity
  - Graph Data

- Copy Number State
- Log2 Ratio
- Weighted Log2 Ratio
- Allele Difference
- B-allele frequency
- Smooth Signal
- Loss of Heterozygosity
- Variants/Somatic Mutations (OncoScan CNV Plus only)

#### ReproSeq aneuploidy (zip files)

When loading zip files from Ion Reporter into ChAS for viewing, the software displays the following segments and graph data:

- Segment data
  - Copy Number Gain/Loss
- •Graph Data
  - Copy Number State (sequence tiles)

### Viewing data and features of interest using the ChAS display controls

ChAS provides the following options for viewing and studying your loaded analysis results data:

• Graphic Displays

See "Displaying Data in Graphic Views" on page 154.

• Tables

See "Displaying data in table views" on page 327.

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" on page 219.

- 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.

See "Using CytoRegions" on page 268.

- Display only segments that appear in CytoRegions using Restricted Mode.
- Query segments from a loaded sample against segments previously uploaded to a ChAS Database.

"Querying a segment from the segment table" on page 390

- See which samples had segments similar to the current sample.
- View the Calls and Interpretations of previous segments to help in the analysis of the current sample.
- 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 and filter" on page 280.

See "Creating and editing AED files" on page 288.

- View genotype data from Next Generation Sequencing data via VCF files.
   See "VCF files" on page 322.
- Prepare reports on your findings by exporting graphics and table data in PDF and other formats.

See "Exporting results" on page 413.

• Save setups of ChAS for different tasks in user profiles and named settings. See "User profiles and named settings" on page 436.

### Working with ChAS

- "Accessing functions in ChAS" on page 42
- "Changing pane sizes" on page 42
- "Opening panes in separate windows" on page 43

### Accessing functions in ChAS

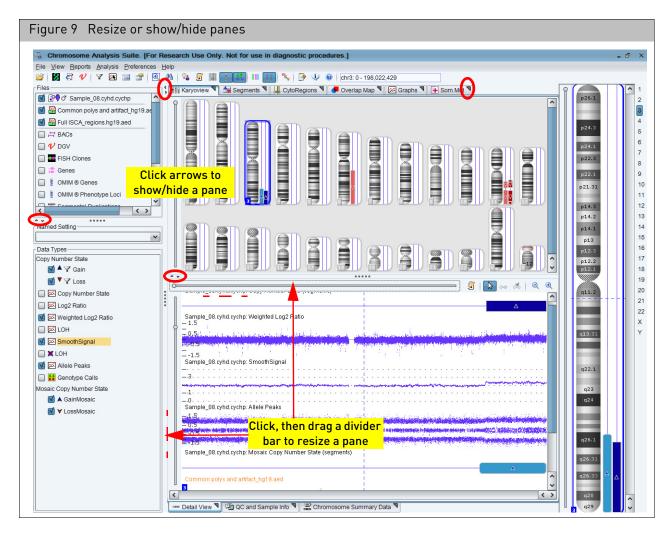
Commands in ChAS can be accessed in the following ways:

- Main menus
- Tool bar
- Right-click menu options in:
  - Files List
  - Data Types List
  - Karyoview
  - Selected Chromosome View
  - Detail View
  - Table Headers

### Changing pane sizes

Do one of the following to change the size of the panes in the ChAS window, as shown in Figure 9 on page 43.

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



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.



# CN/LOH/Mosaicism analysis

#### This chapter includes:

- "Single sample analysis" on page 45
- "Performing a single sample analysis" on page 51
- "Converting CEL files to CHP files" on page 84
- "Reference files" on page 89

ChAS analyzes the intensity data (CEL file) from both CytoScan and OncoScan 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/XNCHP/OSCHP data file that you load and view in ChAS. The analysis detects segments that exhibit are as follows:

- Copy Number State Gain or Loss: Regions of integer copy number gain or integer copy number loss.
- Mosaic Gain and Loss (CytoScan HD, CytoScan 750k and CytoScan Optima only): Regions of non-integer copy number gain or loss (CN states between 1 and 3).
- XON Region Gain and Loss (CytoScan XON only): Regions of gain or loss at the exon level.

**Note:** The mosaicism segmentation analysis is currently only available for CytoScan Array CEL files. However, xxCHP 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 CHP data, see "Loading Data" on page 119.

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

**Note:** Reference files are provided as part of the complete Library file packages. You can also create your own reference file using ChAS.

Load Genome-Wide Human SNP Array 6.0 CNCHP into ChAS to display and detect Copy Number and Loss of Heterozygosity segments. See "Loading Data" on page 119.

Load CytoScan HTCMA Array RHCHP into ChAS to display and detect Copy Number segments, Loss of Heterozygosity segments, variant data, and SMN data. See "Loading Data" on page 119.

**IMPORTANT!** It is recommended to perform analysis operations with all associated analysis files in a locally stored folder(s).

### 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 included default reference file or create your own (For more details, see "Creating a reference file" on page 90)

#### This section includes:

- "ChAS analysis file compatibility"
- "Introduction to single sample analysis" on page 46
- "LOH segments on X and Y chromosomes" on page 49
- "Performing a single sample analysis" on page 51

# ChAS analysis file compatibility

Table 5 lists the compatibility between ChAS Analysis file versions for the CytoScan Arrays.

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

Table 5 Compatibility table

CytoScan Array Analysis File Set Version	ChAS 4.3	ChAS 4.2.1	ChAS 4.2	ChAS 4.1	ChAS 4.0
NA36 (hg38)	Yes	Yes	Yes	Yes	Yes
NA36 (hg19)	Yes	Yes	Yes	Yes	Yes
NA33(hg19)	Yes	Yes	Yes	Yes	Yes
NA32.3(hg19)	No	No	No	No	No
NA32.1(hg19)	No	No	No	No	No
NA32(hg19)	No	No	No	No	No

**Note:** Refer to the ChAS Release Notes for data equivalency information between the ChAS software and the Library file versions used to create CHP files.

# Introduction to single sample analysis

#### Single sample analysis requires:

- ChAS analysis files for the array
   See "Analysis file download" on page 26.
- A previously created reference model file

You can use the included reference model file or create one using your own CEL file data and the Reference File creation function. The Reference Model file in the CytoScan Array set 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:

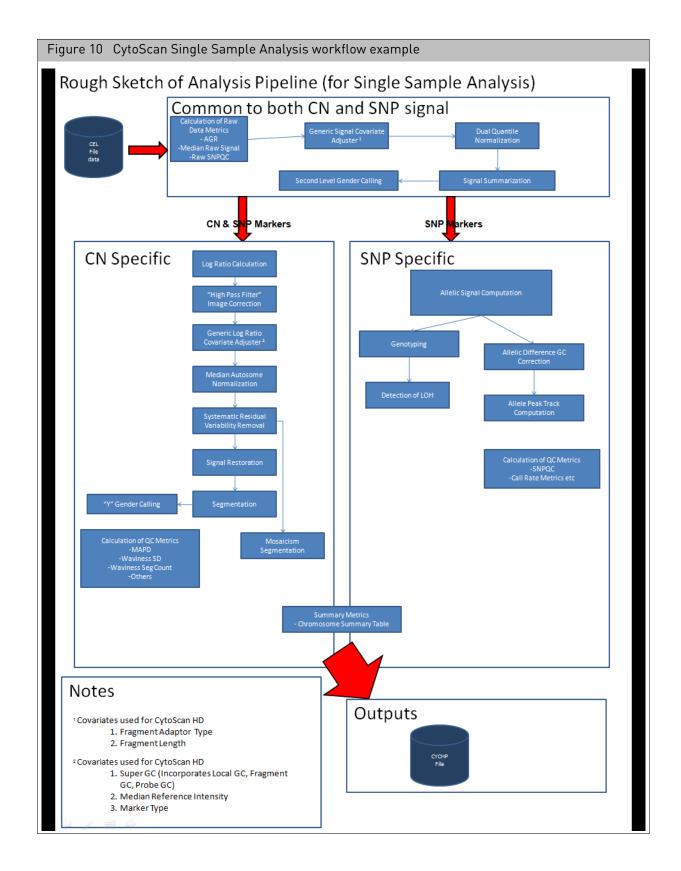
- 284 HapMap samples including at least one replicate of each of 270 HapMap samples: 90 from each of the Yoruban, Asian, and Caucasian ethnic groups, from cell-line derived DNAs from the Coriell Institute of Medical Research.
- 96 DNA samples from blood of phenotypically healthy male and female individuals 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 Difference
  - Smooth Signal
  - Genotype Calls
  - B-allele Frequency
- Segment Data
  - Copy Number Gain/Loss
  - Mosaicism Gain/Loss
  - Loss of Heterozygosity (LOH)

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

The figure below (Figure 10 on page 47) shows an overview of single sample analysis for the CytoScan Array.



#### 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.

#### Mosaic copy number segments on the X chromosome

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.

Mosaic Segments whose boundaries start and end entirely in one of the PAR regions will use CN=2 as normal to determine the type (GainMosaic or LossMosaic) of Mosaic segment to draw.

Because the Mosaicism algorithm can generate segments which cross the PAR boundaries, Mosaic Segments that touch the non-PAR region of the X chromosome use the gender call of the sample to determine the Type of Mosaic segment to draw.

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.

**IMPORTANT!** Mosaic segments are not produced for the Y chromosome because the algorithm only calls segments between CN=1 and CN=3. However, the Allele Difference/Ballele Frequency, and smooth signal tracks can be used to assess mosaicism.

### LOH segments on X and Y chromosomes

#### CytoScan arrays

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.

The tables below briefly describe how the array-specific algorithms call LOH segments for the X or Y chromosome.

Table 6 Expected LOH calls on the X and Y chromosomes for the CytoScan arrays

LOH Segments	X Chromosome	Y Chromosome
Normal male sample (XY)	LOH calls that are single copy-based LOH call (CN = 1).	
Male sample with multiple X chromosomes (for example, XXY)	LOH calls are possible, depending on the constitution of the X chromosomes' origins.	No LOH calls are made for the Y chromosome. Genotype calling is not performed on the Y chromosome.
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 7 Expected LOH calls on the X and Y chromosomes for the Genome-Wide Human SNP Array 6.0

LOH Segments	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.	LOH calls that are due to single copy genotyping calls (CN = 1).
	Heterozygous calls are ignored on the X chromosome in males.	
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.	

Table 7 Expected LOH calls on the X and Y chromosomes for the Genome-Wide Human SNP Array 6.0

LOH Segments	X Chromosome	Y Chromosome	
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	
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.	chromosomal material.	

Table 8 Expected LOH calls on the X and Y chromosomes for the OncoScan arrays

LOH Segments	X Chromosome	Y Chromosome	
Normal male sample (XY)	LOH calls that are single copy-based LOH call (CN = 1).		
Male sample with multiple X chromosomes (for example, XXY)	LOH calls are possible where there is either loss or low heterozygosity.	No LOH calls are made for the Y chromosome.	
Normal female sample (XX)	LOH calls are possible depending on the constitution of the X origins or in regions of either loss or low heterozygosity.		
Female sample with a single X chromosome (X0)	LOH regions on X which have only a single copy. Will be called LOH where there is single copy X.		

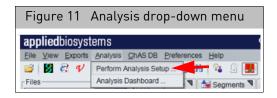
### Performing a single sample analysis

You only need to perform the following steps once, as the data and selections you input (throughout this section) are retained for your convenience with future single sample analysis runs.

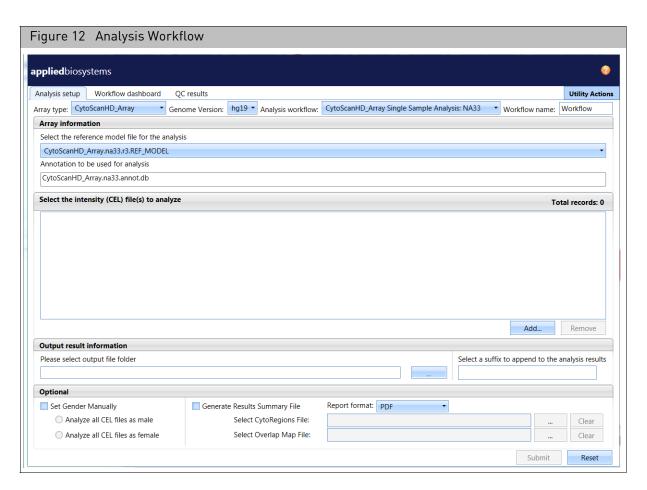
Setting up and running a single sample analysis

**Note:** If you want to setup and run an OncoScan Analysis, see "Setting up and running an OncoScan single sample analysis" on page 67. If your samples are cancer samples and you suspect aberrations for at least 50% of the genome, then running a Normal Diploid Analysis is recommended. For more information, see "Setting up and running a normal diploid analysis" on page 65.

1. From the Analysis menu, select **Perform Analysis Setup.** (Figure 11)



The Analysis Setup window tab opens. (Figure 12)



2. From the **Select array type** drop-down list, click to select CytoScan array type. (Example: CytoScanHD\_Array)

**Note:** Once you select the array type, analysis workflow, and reference model file, then the annotation file will be auto selected for you based on your earlier selections. The Select 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.

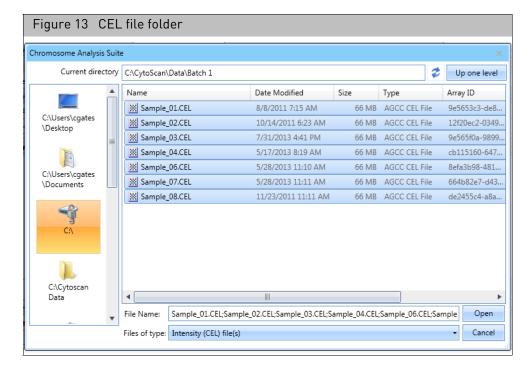
- 3. Choose a Genome Build. (Example: hg38)
- 4. From the **Select analysis workflow** drop-down list, click to select an analysis workflow. (Example CytoScanHD\_Array Single Sample Analysis: NA33 or higher)
- 5. By default, the Set workflow name is **Workflow**. Click inside the Workflow's (upper right) text box to enter a different workflow name.
- From the Select the reference model file for the analysis drop-down list, click to select a reference model file for the analysis. (Example: CytoScanHD\_Array.na33.r2.REF\_MODEL or higher).

**Note:** For Single Sample Analysis, the **Annotation to be used for analysis** field is auto-populated and set with the annotation filename used when the reference model file was generated.

7. At the Select the intensity (CEL) file(s) to analyze pane, click Add

**Note:** The Workflow Analysis window retains the drop-down selections used in your last submitted analysis. However, it does not display a previously set workflow name, CEL files to analyze list, or any suffix used to append your last analysis results. These three fields must be completed again.

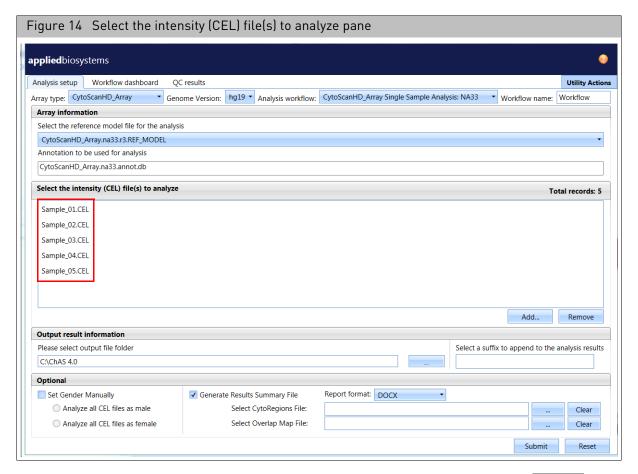
The following window appears: (Figure 13)



- 8. If your CEL files are located somewhere other than your input path location, navigate to the desired folder. Single click, Ctrl click, Shift click or Ctrl a (to select multiple CEL files).
- 9. Click Open.

The **Select the intensity (CEL) file(s) to analyze** pane is now populated with your CEL files, as shown in Figure 14.

**Note:** You can load several CEL files at a time for a Single Sample Analysis.



To remove a CEL file from this list, click to highlight it, then click Remove

10. At the **Output result information** pane, confirm the path shown for your output file folder. To change the current path/folder, click output path/folder.

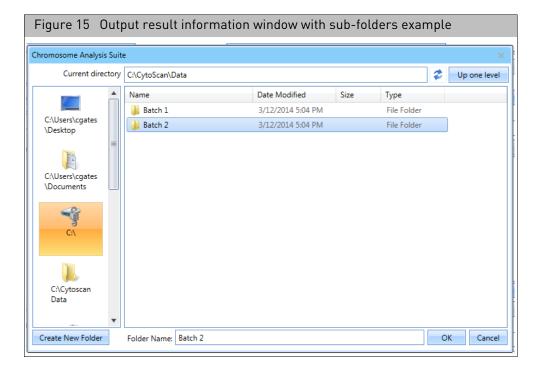
**Note:** To better organize your output results, you can add sub-folders to your assigned output result path/folder.

To add sub-folders to your assigned result path/folder:

- Click to return to your assigned output path and/or folder.
- Click Create New Folder.
- Enter a sub-folder name.
- Click **OK**. Repeat the above steps to add more sub-folders.

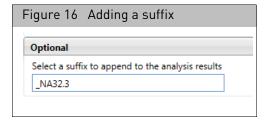
4

The newly created sub-folders now appear in the output result information window. (Figure 15)

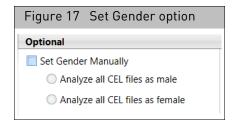


11. If you are using a previously analyzed CEL file(s) to verify new CHP data (against CHP data generated from previous versions of ChAS and Library files), you may want to use a suffix to append the new resulting CHP file(s). To do this, click inside the Select a suffix to append to the analysis results field to enter an appending file suffix. (Figure 16)

**IMPORTANT!** If you are saving the same .CYCHP file into the same output file folder that contains your originally run CYCHP file, a "1" is automatically added into the filename (in addition to any suffix you may add) to differentiate the two runs of identical CEL file names. Example: na33(1).cyhd.cychp

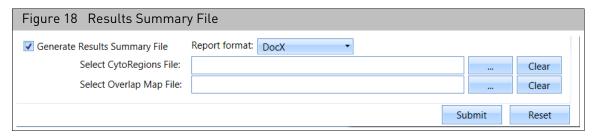


12. Optional: If you have a CEL file(s) in which the Y chromosome is partially/fully deleted and therefore determined to be female by the gender calling algorithm, go to the Analysis Setup's Optional pane (Figure 17), click the **Set Gender Manually** check box, then click to select the appropriate radio button.



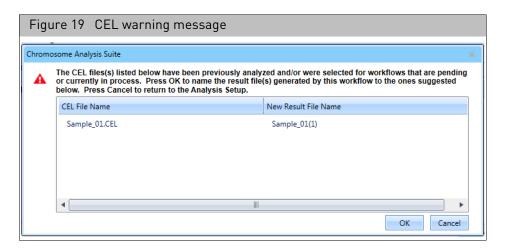
13. Optional: If you want to have an automatic export of the Karyoview, Segments Table, and Detail View for Copy Number and LOH Segments in the CHP file, click the check box **Generate a Results Summary File**, then and select the output format of either PDF or DOCX. (Figure 18)

**Note:** You can assign a CytoRegion and Overlap Map region file that will highlight these regions in the export. The export is placed in the same folder as the CYCHP file. This automatic export feature is only available for CytoScan arrays.



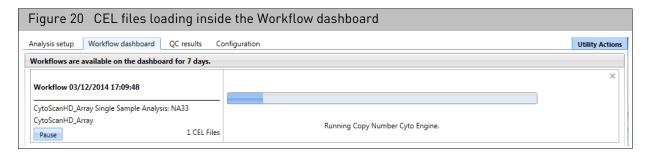
14. Click Submit .

If the following warning message appears (Figure 19), acknowledge it, then click **OK**.

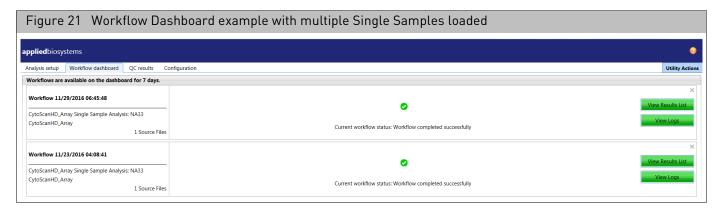


The Workflow dashboard window appears and your annotation files begin to load. (Figure 20).

The Analysis Workflow Dashboard tracks ongoing analysis tasks for ChAS. It also delivers the results of the analyses and can restart the Browser (if it was shut down to free up memory for the analysis).



After loading is complete, a Workflow completed successfully message appears. (Figure 21)

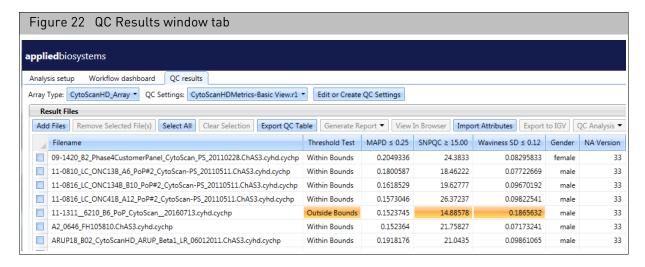


**Note:** The View Logs button will access the algorithm pipeline logs which may be useful if you have a Workflow that fails to complete.

15. Click to choose the analysis you want to view, then click View Results List

The QC Results tab window appears showing the Basic View QC settings. A Detail View QC setting, which provides more columns of data, is also available in the QC Settings drop-down list. (Figure 22)

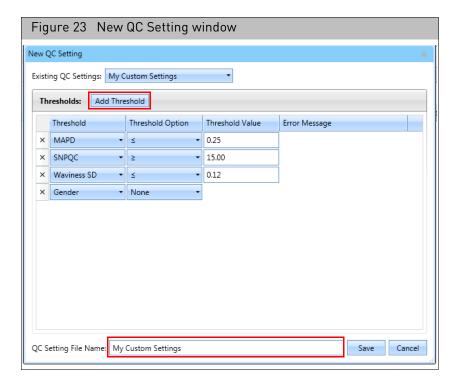
**Note:** QC parameters can also be viewed in the ChAS Browser see setting QC parameters in ChAS Browser.



16. Click each sample's check box or click the 'Select All' button to select all samples.

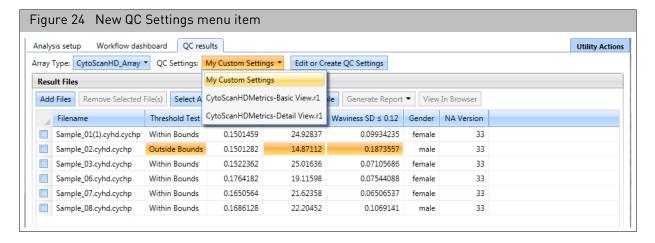
To create your own custom QC setting:

Click on the Edit or Create QC Settings button.
 The New QC Setting window appears. (Figure 23)



- 2. Click **Add Threshold** (Figure 23) to create a new row.
- Select the threshold you want to view in your custom QC Setting.
   Note: The threshold metric you select is not required to have a threshold value.
- 4. Name your custom QC settings. (Example: My Custom Settings) (Figure 23), then click **Save**.

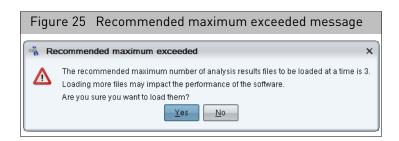
Your custom QC Setting is now available from the QC Settings drop down menu, as shown in Figure 24.



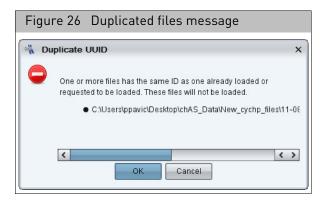
#### To view results in the browser:

At the QC Results window, click the **View in Browser** button or the **View in MSV** button. For more MSV information, see the RHAS User Guide.

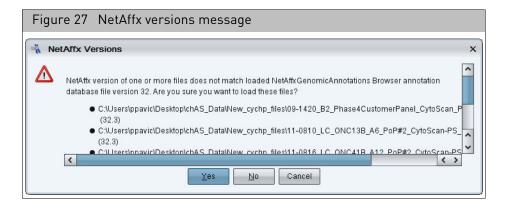
If the following warning message appears (Figure 25), acknowledge it, then click **OK**.



If the following warning message appears (Figure 26), acknowledge it, then click **OK**.



If the following warning message appears (Figure 27), click **Yes** to acknowledge it.

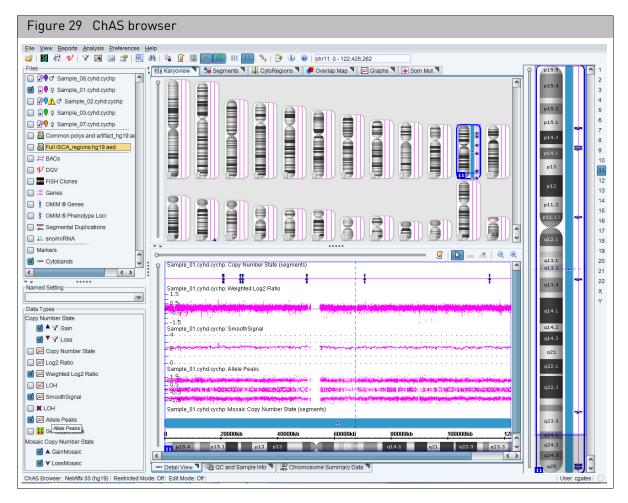


A progress bar appears. (Figure 28)



**Note:** The ChAS Browser allows for loading of xxCHP files analyzed from different versions of ChAS. However, xxCHP files analyzed from different genome versions (hg18, hg19, hg38) cannot be loaded at the same time.

After a few moments, the ChAS browser featuring your selected samples appears. (Figure 29)

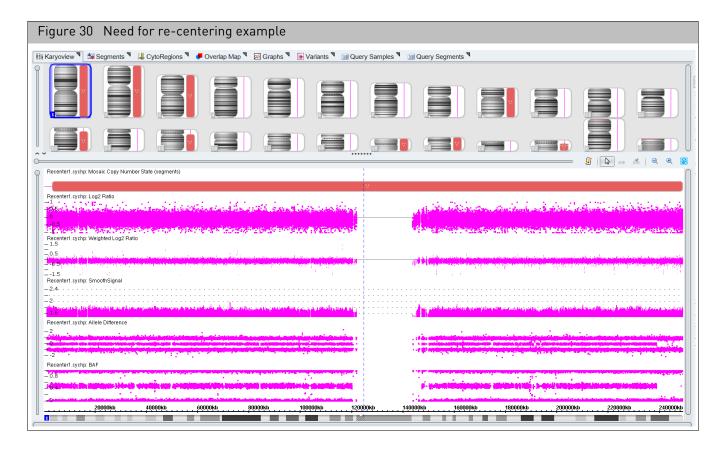


Recentering CytoScan HD, 750K and Optima arrays

Due to the complexity and low diploid count in a small fraction of cancer samples, there may be a need to manually assign the diploid region of the sample or recenter it.

In Figure 30, Chromosome 1 is called as a mosaic copy number loss, the log 2 ratio data is shifted downward, the smooth signal averages 1.75 copies, but the Allele Difference (AD) and B- Allele Frequency (BAF) Graphs are displaying three tracks.

**Note:** Since it is unlikely to have three tracks in AD/BAF data in a region of loss (unless the loss is CN=0), this sample needs to be recentered.

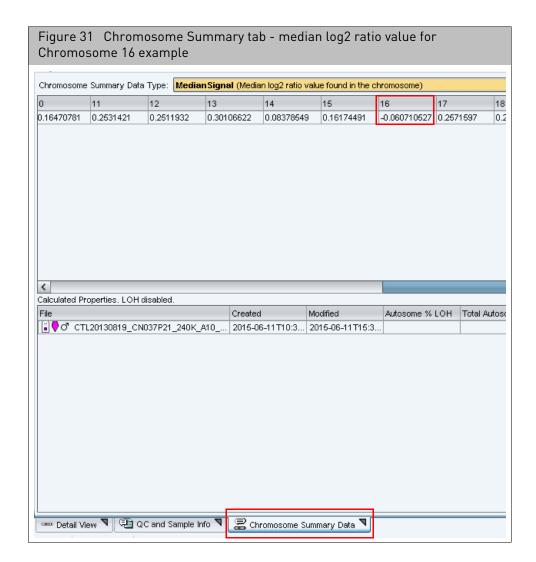


If the region that is true diploid is a whole chromosome, use **Method 1**. If the region that is true diploid is part of a chromosome, use **Method 2**.

#### Method 1

To determine the median Log 2 ratio for the region in the sample that is truly diploid:

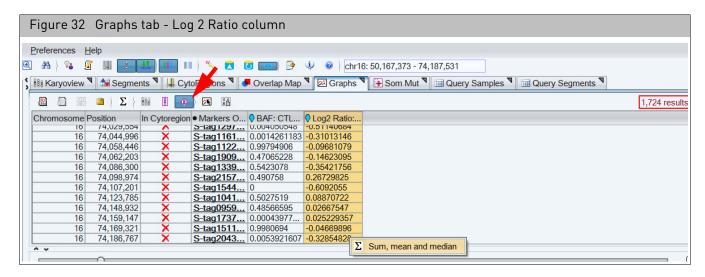
- 1. Open the ChAS Browser.
- 2. Click on the **Chromosome Summary Data** tab, then click the drop-down to select **MedianSignal**. (Figure 31)



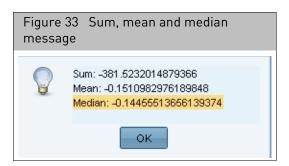
#### Method 2

To determine the median Log 2 ratio for the region in the sample that is truly diploid:

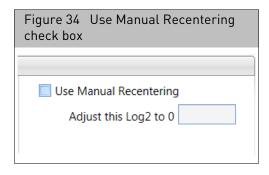
- 1. Open the ChAS Browser.
- 2. In the Detail View (Figure 30), zoom into the region determined to be diploid.
- 3. Go to the **Graphs** tab, then click to include only the selected view.
- 4. Highlight the **Log2 Ratio** column, then right-click to select **Sum**, **mean and median**. (Figure 32)



A Sum, mean and median window appears. (Figure 33)



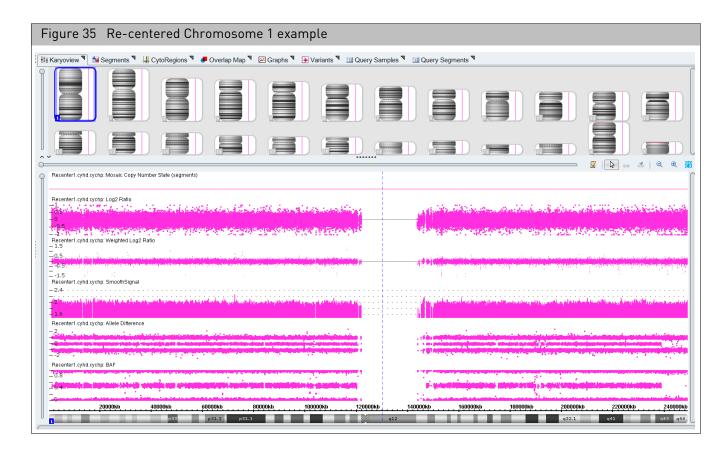
- 5. Acknowledge the message, then click **OK**.
- 6. In the Analysis Workflow, set up the CEL file in the Single Sample Workflow, as described in "Performing a single sample analysis" on page 51.
- 7. Check the Use Manual Recentering check box to enable the parameter fields, then enter the value of the median Log2 (determined by the browser) into the Adjust this log 2 to 0 text field. (Optional) Enter a suffix. Note: A suffix is recommended in order to differentiate the re-centered CYCHP file from the original.



8. Click Submit.

4

**Note:** By entering a median Log2 Ratio value (for the region you have determined to be diploid, The Recentering Algorithm has re-centered the log2 ratio data (for the region determined to be diploid) around 0 and there is no longer a mosaic loss segment called in this region, as shown in the Chromosome 1 example below. (Figure 35)

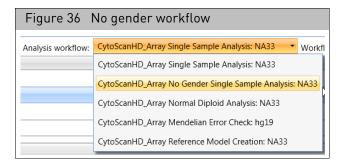


# No gender single sample analysis

The No Gender Single Sample Analysis (Figure 36) is the same analysis as described previously for Single Sample Analysis with the exception that no gender information is displayed in ChAS. The gender will not be reported and no segment or probe level data from X or Y chromosomes are displayed.

The metric, **Sex Chromosomes Aberrated** can be added to the QC table and reports either a Yes or No.

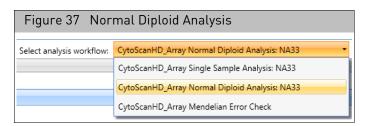
- Yes: Indicates that the sample does have segments meeting the following default thresholds: 50 Markers/200kb for copy number and 50 Markers/ 10,000kb for LOH segments.
- No: Indicates no copy number or LOH segments meet the previously defined thresholds.

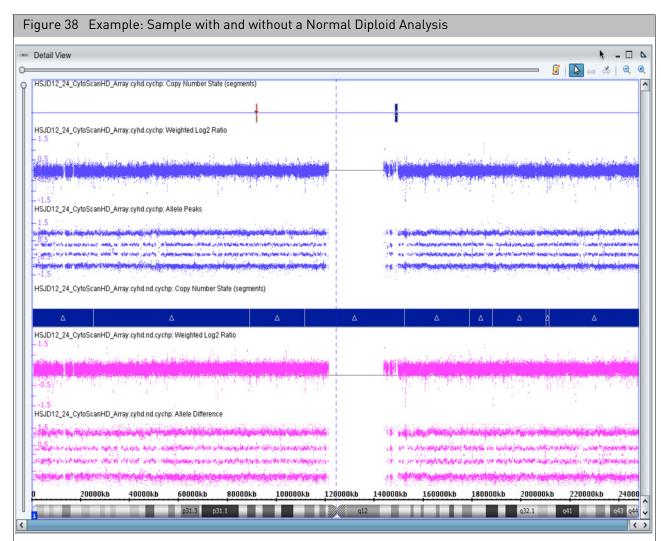


#### Setting up and running a normal diploid analysis

The Normal Diploid Analysis for CytoScan is recommended for cancer samples in which >50% of the genome is likely to be rearranged. This analysis will automatically determine the normal diploid regions and normalize the rest of the sample based on those regions resulting in properly centered data.

A Normal Diploid Analysis has the identical setup steps as "Setting up and running a single sample analysis" on page 51. The only difference is you must select **Normal Diploid Analysis** from the **Select analysis workflow** drop-down menu, as shown in Figure 37.





The data shown above is from a cancer sample in which >50% of the genome was non-diploid. The top graph (purple) shows the sample run through the traditional single sample analysis. There are no Copy Number Segments called, the weighted log2 is centered around 0, but there are 4 allele difference tracks indicating more than two copies of this chromosome. In the bottom graph (pink), this same sample is run through the Normal Diploid normalization algorithm. The Copy Number Gain segment is called, the weighted log2 is shifted above the 0 line which is in agreement with the four allele difference tracks.

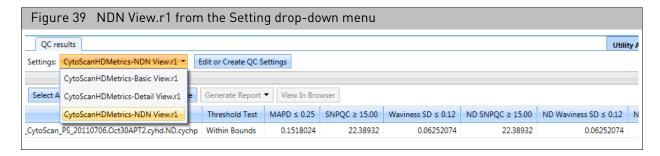
#### The recommended QC metrics for Normal Diploid Analysis are:

- ndSNPOC
- MAPD
- ndwavinessSD
- SNPQC
- wavinessSD

#### Do the following to view the recommended QC metrics (listed above):

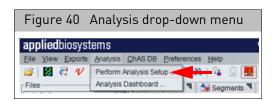
**Note:** For samples run through the Normal Diploid Analysis, the following QC metrics are recommended:

- -MAPD < 0.25
- SNPQC or ndSNPQC >= 15
- wavinessSD or ndwavinessSD < 0.12
- 1. From the Analysis Workflow, click the **QC Results** window tab.
- 2. Click the **Settings** drop-down menu, then select **NDN View.r1**, as shown in Figure 39.

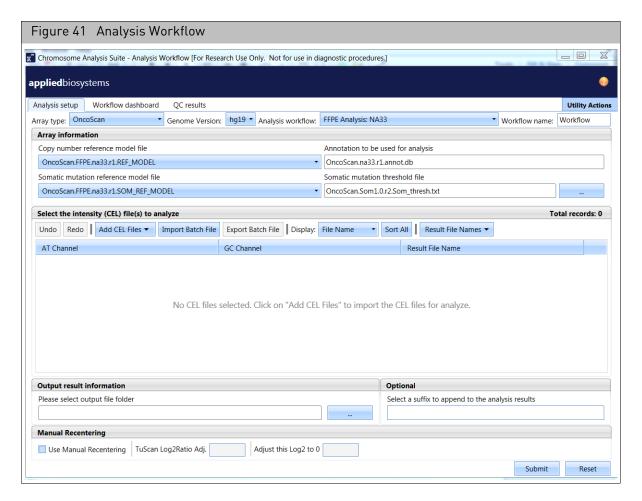


Setting up and running an OncoScan single sample analysis

1. From the Analysis menu, select **Perform Analysis Setup.** (Figure 40)



The Analysis Setup window tab opens (Figure 41).



2. From the **Select array type** drop-down list, click to select **OncoScan**.

**Note:** Once you select the array type, analysis workflow, and reference model file, then the annotation file will be auto selected for you based on your earlier selections.

**IMPORTANT!** The Select 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.

- 3. Select the appropriate Genome Build.
- 4. From the **Select analysis workflow** drop-down list, click to select an appropriate analysis workflow.

**IMPORTANT!** For FFPE samples use the FFPE Analysis NAXX workflow. For Control DNA use the Control DNA Analysis.

- 5. By default, the **Set workflow name** is **Workflow**. Click inside the Workflow's (upper right) text box to enter a different workflow name.
- 6. Click the **Copy number reference model file** from the drop-down list, then click to select an appropriate file.
- 7. Click the **Somatic mutation reference model file** from the drop-down list, then click to select an appropriate file.

- 8. At the **Select the intensity (CEL) file(s) to analyze** pane, click the **Add CEL Files** drop-down, then click to select **AT Channel**.
  - An Explorer Window appears.
- 9. Highlight the CEL file(s) using Ctrl click or Shift click, then click **Open**.
- 10. At the **Select the intensity (CEL) file(s) to analyze** pane, click the **Add CEL Files** drop-down, then click to select **GC Channel**.
  - An Explorer Window appears.
- 11. Highlight the CEL file(s) using Ctrl click or Shift click, then click **Open**.

The Select the intensity (CEL) file(s) to analyze pane is now populated with AT and GC Channel CEL files.

**IMPORTANT!** After loading the CEL files, check that the AT lines up with the matching GC CEL file.

12. Click the **Result File Names** drop-down menu to enable ChAS to automatically generate Output Names.

**Note:** Output file names are only auto-generated if the two CEL files have the same root name. It is recommended to use an "A" or "C" as the last character to designate the channel in the CEL file naming convention. Example: "\_AS\_05A.CEL" is an AT Channel file, while "\_AS\_05C.CEL" is a GC Channel

You can also clear this (populated) column by clicking **Clear Column**.

- 13. OPTIONAL: To choose a different output folder from the saved output path that is displayed, click the Output result information's **Browse** button.
  - An Explorer window appears.
- 14. Navigate to an output folder location, then click **OK**.

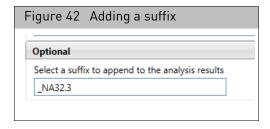
**Note:** To better organize your output results, you can add sub-folders to your assigned output result path/folder.

#### To add sub-folders to your assigned Output Results folder:

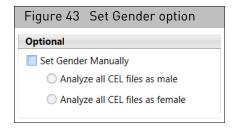
- Click to return to your assigned output path and/or folder.
- Click Create New Folder.
- Enter a sub-folder name.
- Click **OK**. Repeat the above steps to add more sub-folders.

  The newly created sub-folders now appear in the output result information
- 15. If you are using a previously analyzed CEL file(s) to verify new CHP data (against CHP data generated from previous versions of ChAS and Library files), you may want to use a suffix to append the new resulting CHP file(s). To do this, click inside the **Select a suffix to append to the analysis results** field to enter an appending file suffix. (Figure 42)

**IMPORTANT!** If you are saving the same OSCHP file into the same output file folder that contains your originally run OSCHP file, a "1" is automatically added into the filename (in addition to any suffix you may add) to differentiate the two runs of identical CEL file names. Example: na33(1).oschp

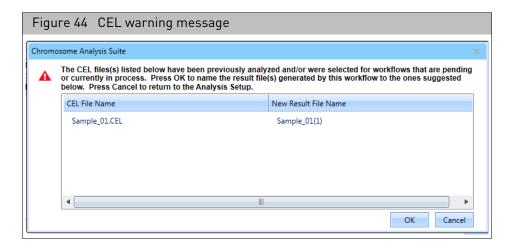


16. Optional: If you have a CEL file(s) in which the Y chromosome is partially/fully deleted and therefore determined to be female by the gender calling algorithm, go to the Analysis Setup's Optional pane (Figure 43), click the **Set Gender Manually** check box, then click to select the appropriate radio button.



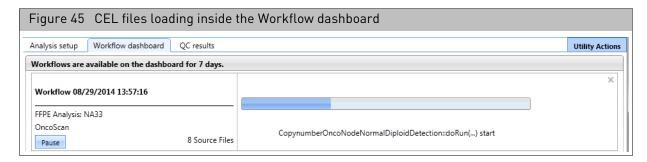
17. Click Submit

If the following warning message appears (Figure 44), acknowledge it, then click **OK**.



The Workflow dashboard window appears and your annotation files begin to load. (Figure 45).

The Analysis Workflow Dashboard tracks ongoing analysis tasks for ChAS. It also delivers the results of the analyses and can restart the Browser (if it was shut down to free up memory for the analysis).



After loading is complete, a Workflow completed successfully message appears. (Figure 46)

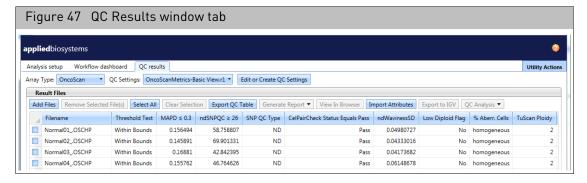


**Note:** The View Logs button will access the algorithm pipeline logs which may be useful if you have a Workflow that fails to complete.

18. Click to choose the analysis you want to view, then click **View Results List**.

The QC Results tab window appears showing the Basic View QC settings. (Figure 47) A Detail View QC setting (which provides more columns of data) is also available in the QC settings drop down list.

**Note:** QC parameters can also be viewed in the ChAS Browser see "Setting QC parameters in the ChAS browser" on page 134.



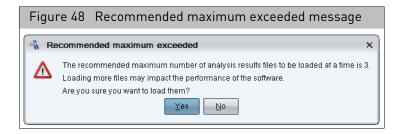
19. Click each sample's check box or click the **Sample File** check box to select ALL samples.

#### Creating your own custom QC settings

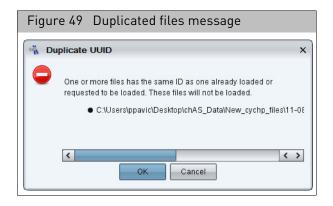
See "To create your own custom QC setting:" on page 57

#### To view results in the Browser:

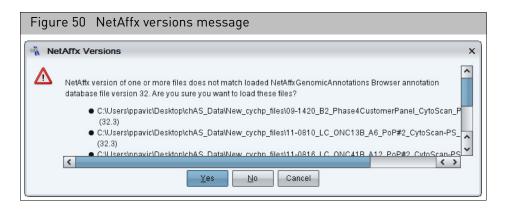
At the QC Results window, click View In Browser.
 If the following warning message appears (Figure 48), click Yes to acknowledge it.



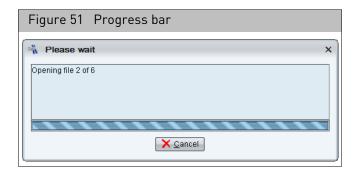
If the following warning message appears (Figure 49), acknowledge it, then click **OK**.



If the following warning message appears (Figure 50), click **Yes** to acknowledge it.



A progress bar appears. (Figure 51)

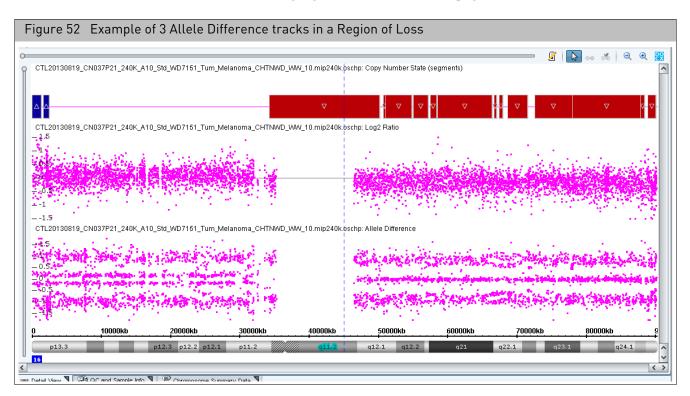


**Note:** The ChAS Browser allows loading files analyzed using different NetAffx version at the same time (as long as the versions are all from all the same reference and genome builds). If NetAffx versions are from different builds of the genome (for example Hg18 and Hg19), The ChAS Browser does not load the files.

After a few moments, the ChAS browser featuring your selected samples appears.

Recentering OncoScan CNV and OncoScan CNV Plus arrays Due to the complexity and low diploid count in a small fraction of cancer samples, there may be a need to manually assign the diploid region of the sample or "recenter" it.

In Figure 52, Chromosome 16q is called as a loss, the log 2 ratio data is shifted downward, but the Allele Difference Graph is displaying three tracks representing AA, AB, BB calls. Having an Allele Difference graph with three tracks means this region must have at least two copies. Since you cannot have three Allele Difference tracks in a region of loss, this sample needs to be recentered. For more information, see "Manual re-centering algorithm (OncoScan)" on page 506.

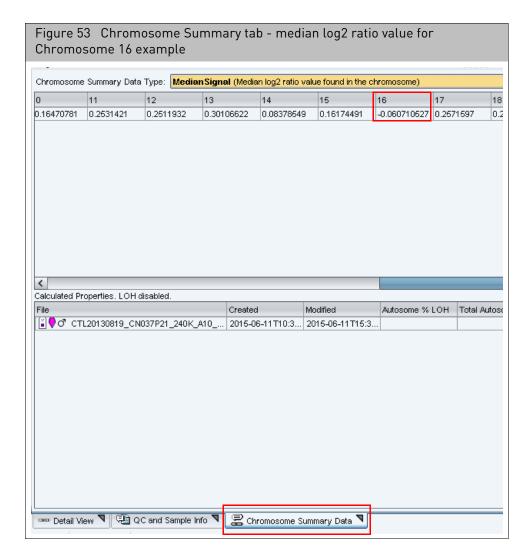


If the region that is true diploid is a whole chromosome, use **Method 1**. If the region that is true diploid is part of a chromosome, use **Method 2**.

#### Method 1

To determine the median Log 2 ratio for the region in the sample that is truly diploid:

- 1. Open the ChAS Browser.
- 2. Click on the **Chromosome Summary Data** tab, then click the drop-down to select **MedianSignal**. (Figure 53)

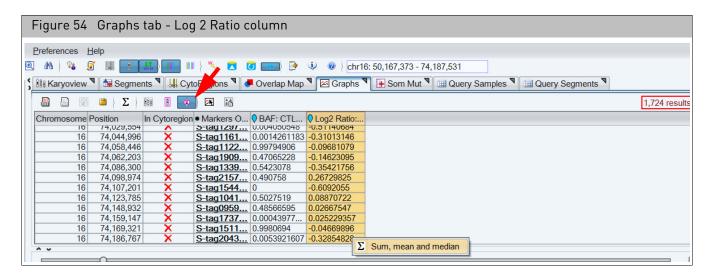


#### Method 2

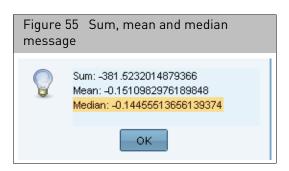
To determine the median Log 2 ratio for the region in the sample that is truly diploid:

- 1. Open the ChAS Browser.
- 2. In the Detail View (Figure 52), zoom into the region determined to be diploid.
- 3. Go to the **Graphs** tab, then click to include only the selected view.

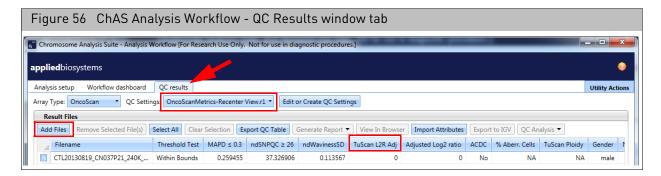
4. Highlight the **Log2 Ratio** column, then right-click to select **Sum**, **mean and median**. (Figure 54)



A Sum, mean and median window appears. (Figure 55)

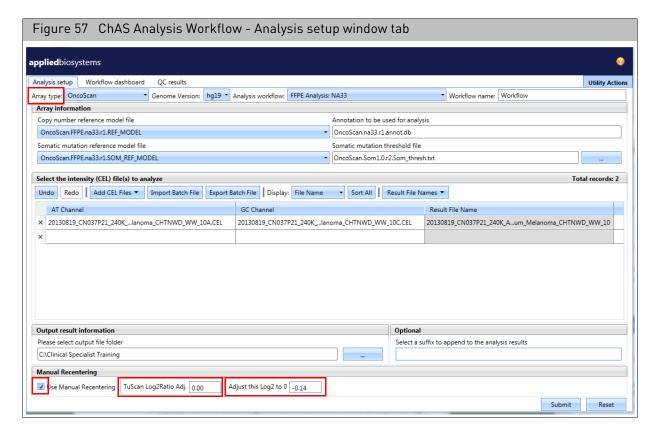


- 5. Acknowledge the message, then click **OK**.
- 6. Open the **ChAS Analysis Workflow**, then click on the **QC Results** tab. (Figure 56)



- 7. Load the OSCHP file into the **QC Results** tab by clicking on the **Add Files** button. (Figure 56)
- 8. From the QC Settings drop-down menu, select Recenter View, then make a note of the TuScan L2R Adj value. (Figure 56)

### 9. Click on the Analysis setup tab. (Figure 57)

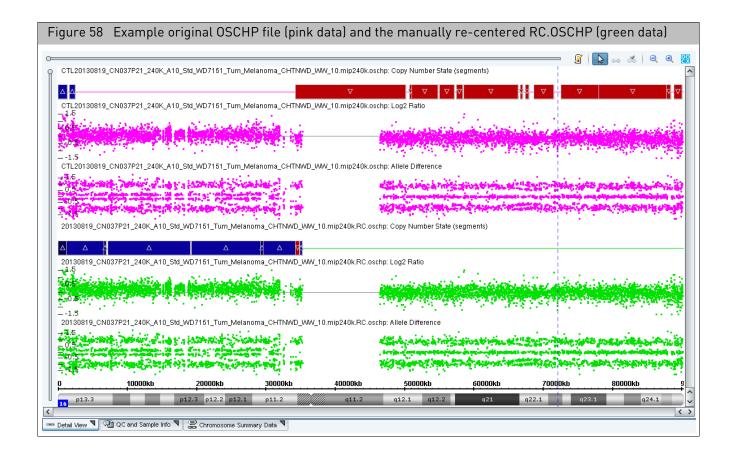


- 10. Select the FFPE or non-FFPE analysis based on the sample type.
- 11. Load in the two CEL files into the appropriate channel.
- 12. Check the Manual Recenter check box to enable the parameter fields. (Figure 57)
- 13. Enter the **TuScan Log2 Adjustment** value you noted earlier.
- 14. Enter the value of the **median Log2** determined from the browser into the **Adjust** this log 2 to 0 field.
- 15. Enter a suffix if desired.

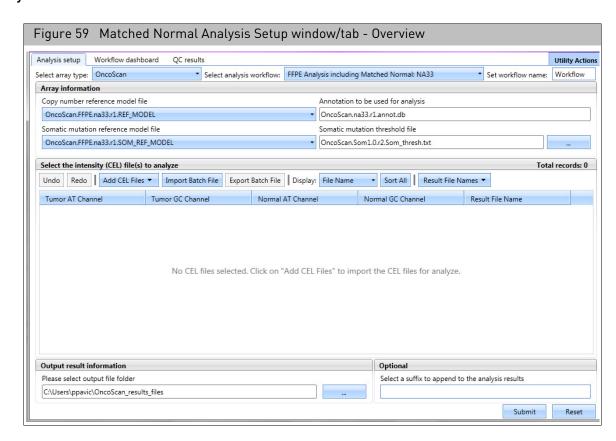
**Note:** An RC will automatically be appended onto any OSCHP file that goes through Manual Recentering for an RC.OSCHP extension.

Figure 58 shows the original OSCHP file (pink data) and the manually recentered RC.OSCHP (green data).

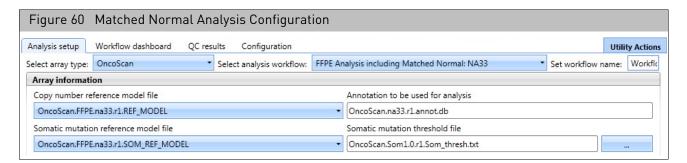
By inputting both the TuScan Log2 Ratio value (derived from the algorithm) and the median Log2 Ratio value (for the region you have determined to be diploid, Chromosome 16q for our example), the Recentering Algorithm has recentered the log2 ratio data (for the region determined to be diploid) around 0 and there is no longer a loss segment called in this region.



Setting up and running an OncoScan matched normal analysis



As long as your library file folder contains the necessary analysis files for the array, your configuration paths are established and your Array Information fields autopopulate. (Figure 60)



1. From the Select array type drop-down list, select OncoScan.

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

**IMPORTANT!** After adding new library files to the library file folder, always close and relaunch OncoScan Console to ensure the newly added files are recognized by the software.

From the Select analysis workflow drop-down list, click to select FFPE Analysis including Matched Normal NAXX.

Other available Analysis Workflow options are:

- Control DNA Analysis NAXX Use this workflow for the Control DNA in the OncoScan Kit.
- Non-FFPE Analysis NAXX Use this workflow with cell line DNA.
- FFPE Analysis NAXX Use this workflow for a standard analysis.
- 3. Enter a Workflow name (optional). By default, the Set workflow name is **Workflow**. Click workflow (upper right) to enter a different workflow name.

**Note:** Customizing a Workflow name can be a useful tool in keeping track of analysis workflows as all the related output files (outside of the OSCHP file) are pre-fixed with this workflow name.

The Annotation file is automatically selected for you and is based on your selected reference model file. (Example: **OncoScan.na33.v1.annot.db**)

**Note:** The Annotation to be used for analysis field is auto-populated based on your Ref Model file selection. The analysis is not be permitted to run if the appropriate annotation file is not available in your Library folder.

- 4. Select a Somatic mutation reference model file. By default, it is set to **OncoScan.na33.v1.SOM\_REF\_MODEL**. If you created your own reference model file, click the drop-down list to select your .SOM\_REF\_MODEL.
- 5. Confirm the displayed Somatic mutation threshold file to be used is correct. If you need to change it, click the **Browse** button, navigate to the appropriate threshold TXT file, then click **OK**.

**IMPORTANT!** If the Reference Model File and Somatic mutation Reference Model File were created independently of each other, a warning message appears after you click Submit (to start the Workflow Analysis process). Click OK to acknowledge the message.

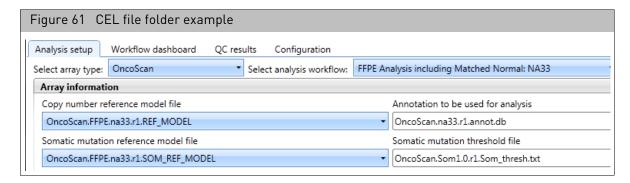
# Adding CEL files to analyze

You can manually add CEL files or import them as a tab-delimited text file.

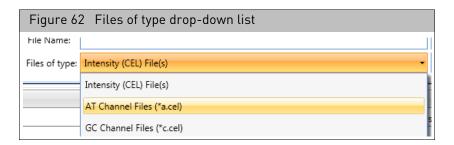
#### Manually adding CEL files to analyze

- 1. At the **Select the intensity (CEL) file(s) to analyze** pane, click the **Add CEL files** drop-down.
- 2. Click Tumor AT Channel.

The CEL file window appears. (Figure 61)



3. Click any header to sort your files or click the **Files of type** drop-down to filter your CEL files by AT Channel, as shown in Figure 62.

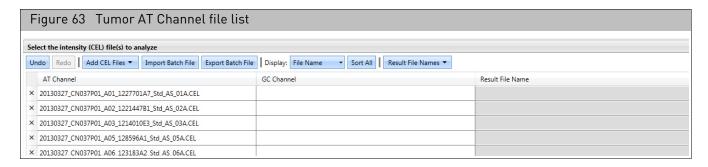


4. Single click, Ctrl click, or Shift click (to select multiple Tumor AT Channel files).

**IMPORTANT!** It is recommended to use an "A" or "C" as the last character to designate the channel in the CEL file naming convention. Example: "\_AS\_05A.CEL" is an AT Channel file, while "\_AS\_05C.CEL" is a GC Channel file. See Figure 61.

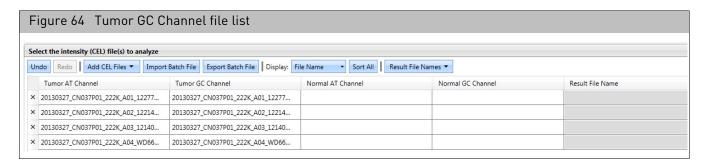
5. Click **Open**.

The Tumor AT Channel fields are now populated. (Figure 63)



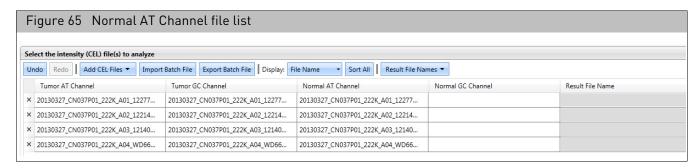
- 6. Click the **Add CEL files** drop-down.
- 7. Click **Tumor GC Channel**. The CEL file window appears. (Figure 61 on page 80)
- 8. Single click, Ctrl click, or Shift click (to select multiple Tumor GC Channel files).
- 9. Click Open.

The Tumor GC Channel fields are now populated. (Figure 64)



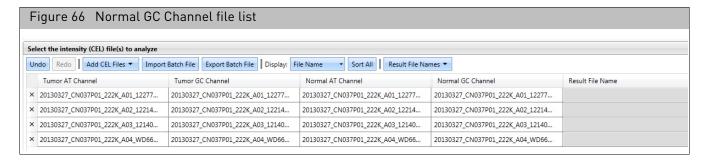
- 10. Click the **Add CEL files** drop-down.
- 11. Click Normal AT Channel. The CEL file window appears. (Figure 61)
- 12. Single click, Ctrl click, or Shift click (to select multiple Normal AT Channel files).
- 13. Click Open.

The Normal AT Channel fields are now populated. (Figure 65)



- 14. Click the Add CEL files drop-down.
- 15. Click **Normal GC Channel**. The CEL file window appears. (Figure 61)
- 16. Single click, Ctrl click, or Shift click (to select multiple Normal GC Channel files).
- 17. Click Open.

The Normal GC Channel fields are now populated. (Figure 66)

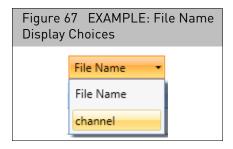


## CEL file displaying options (optional)

The File Name drop-down list (Figure 67) is dynamically populated and based on what attributes are populated in the ARR file.

### To use this display option, you must:

- 1. Provide the appropriate attributes at the time of sample registration in AGCC.
- 2. The ARR files must reside in the same folder as the CEL files.



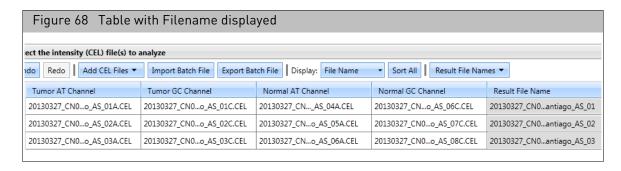
To see "channel" (as an option in the drop down), you must use a template (or the OncoScan template provided in the AGCC library files) that contains a "channel" attribute. The resulting ARR file must also reside in the same folder as the CEL files you are analyzing.

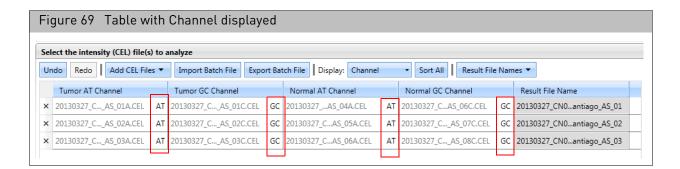
You can display one of the attributes from the ARR file in the table. For example, "Channel" can be chosen (Figure 67) to confirm the assignment of a CEL file to its appropriate channel.

### To select a File Name display attribute:

1. Click the **File Name** drop-down button, then click to select the attribute you want displayed along with your CEL file names.

The two examples (Figure 68 and Figure 69) show how the table appears with the display set to Filename, then to Channel.





# Importing CEL files using batch import

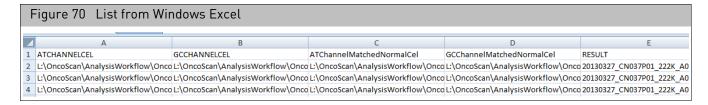
The batch file must be saved as a text (Tab-delimited) format and include the full directory path for your CEL files, as shown in Figure 70.

**Note:** The resulting OSCHP files are saved to your output path location, therefore it is not necessary to include a path under RESULT. Simply enter the desired results filename in this column.

The format for this tab-delimited file is 5 columns (A,B, C, D, and E) with the headers:

- ATCHANNELCEL
- GCCHANNELCEL
- ATChannelMatchedNormalCel
- GCChannelMatchedNormalCel
- RESULT

You must provide the full path to the CEL files for each Channel column. (Example: C:\Desktop\OncoScan\Data\Sample1.cel)



## 1. Click Import Batch File

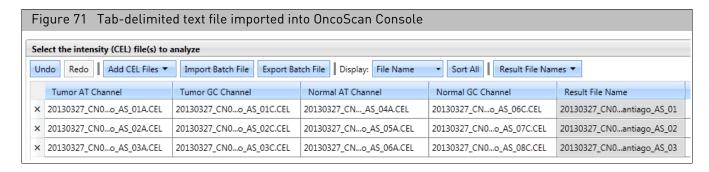
A File window appears.

2. Navigate to your text (tab delimited) file location, then click on the file you want to import.

### **IMPORTANT!** Excel must be closed before you import.

### 3. Click Open.

The Tumor AT, Tumor GC, Normal AT, Normal GC and Result File Name fields are now populated. (Figure 71)



# Converting CEL files to CHP files

The **Automatic CEL File Analysis** tool in ChAS converts CEL files into ChAS Browser compatible CHP files automatically.

#### How it works

The Automatic CEL File Analysis tool continually scans up to five designated input folders for new CEL files to analyze. Once a CEL file is detected, it is analyzed resulting in a CHP file (which is auto-saved to a designated output folder). Once a CEL file is processed, the ARR and CEL files are moved to the designated Archive folder.

# Supported array types

- CytoScan HD (Single Sample Analysis or Normal Diploid Analysis only)
- CytoScan XON (Single Sample Analysis only)
- CytoScan 750K (Single Sample Analysis or Normal Diploid Analysis only)
- CytoScan Optima (Single Sample Analysis only)
- OncoScan CNV (FFPE Analysis only)
- OncoScan CNV Plus (FFPE Analysis only)

**Note:** In order to properly match up CEL files when running OncoScan CNV or OncoScan CNV Plus arrays, the CEL files from the "AT" and "GC" channels must have the exact same root name and include the array type. Each CEL file must indicate which channel they are using a concluding A or C. For example, the following CEL files will be paired properly for OncoScan FFPE analysis:

- OncoScan\_Sample1\_(OncoScanCNV\_Array)A.cel
- OncoScan\_Sample1\_(OncoScanCNV\_Array)C.cel

The following CEL files will NOT be paired properly for OncoScan FFPE analysis because the root names do not match:

- OncoScan\_Sample1\_(OncoScanCNV\_Array)A.cel
- OncoScan\_Sample2\_(OncoScanCNV\_Array)C.cel

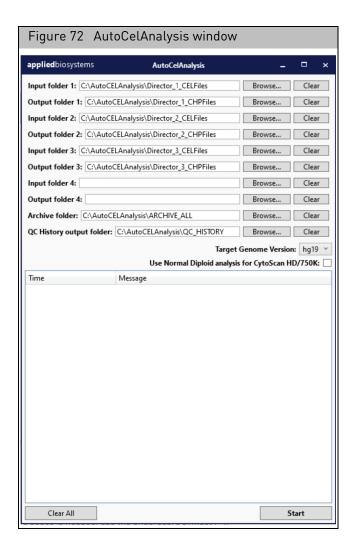
The following CEL files will NOT be paired properly for OncoScan FFPE analysis because the array names are not included:

- OncoScan\_Sample1\_A.cel
- OncoScan\_Sample1\_C.cel

## Launching the tool

- 1. Click the **Utility Actions** button (top right of the Analysis Workflow window)
- 2. Click Automatic CEL File Analysis.

The window opens. (Figure 72)



# Setting up the tool

**IMPORTANT!** Input, Output, and Archive folder names and their assigned paths should not contain spaces. If a space is needed, use the underscore symbol [\_].

Any non-ARR and non-CEL files detected in your assigned input folder will remain in this folder. Any detected ARR and CEL files will be processed and moved into your assigned Archive folder.

## Assigning your input folder(s)

The Input folder is where your CEL files reside.

- Click the Input folder 1 Browse... button.
   An Explorer window appears.
- 2. Navigate to your CEL file folder.
- 3. Click on the folder to highlight it, then click **Select Folder**.

Your CEL file folder path is displayed.

4. Optional: Once the first Input folder is assigned, an option to designate a second Input folder appears (Input folder 2). Repeat steps above to assign a second Input folder. Up to five Input folders can be assigned.

# Assigning your output folder(s)

After the tool produces a CHP file, it is placed in an Output folder.

- Click the Output folder 1 Browse... button.
   An Explorer window appears.
- 2. Navigate to an existing folder where you want your newly converted CHP files saved or click **New Folder** to create a new Output folder.
- 3. Click Select Folder.

Your Output folder 1 path is displayed.

4. Optional: Once the first Output folder is assigned, an option to designate a second Output folder appears (Input folder 2). Repeat steps above to assign a second Output folder. Up to five Output folders can be assigned.

## Assigning an archive folder

Copies of the processed CEL files are placed in the Archive folder.

- 1. Click the **Archive folder** Browse... button.
  - An Explorer window appears.
- 2. Navigate to the folder where you want the CEL files saved or click **New Folder** to create a new Archive folder.
- 3. Click Select Folder.

Your Archive folder path is displayed.

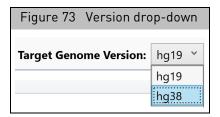
#### Assigning a QC History file folder

- 1. Click the **QC History folder** button.
  - An Explorer window appears.
- 2. Navigate to the folder where you want the CEL files saved or click **New Folder** to create a new QC History folder.
- 3. Click **Select Folder**.

Your QC History folder path is displayed.

# Selecting a target genome version

1. By default, **hg19** is selected. If needed, click the **Target Genome Version** drop-down to select **hg38**, a shown in Figure 73.



## Normal diploid analysis check box

Use Normal Diploid analysis for CytoScan HD/750K:

1. Click the check box to use the Normal Diploid algorithm. Leave unchecked to run the default single sample analysis.

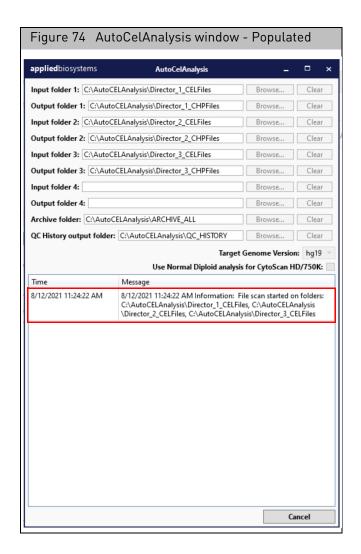
# Running the tool

1. Click Start

The CEL file analysis begins and the Time and Message pane populates with information, as shown in Figure 74. Click **Cancel** to stop an analysis in progress. **Note:** The time for CEL file analysis to complete should be comparable to analysis times when manually setting up CEL files in the Analysis Workflow.

**Note:** During the analysis process, the tool detects the array type, then autogenerates an appropriately named array staging folder. As each CEL file analysis completes, the tool moves the newly created CHP file from its staging folder to the output folder you assigned earlier.

**Note:** If the Automatic CEL File Analysis stops for any reason (such as losing connection to the input folder or is inadvertently canceled) a log file of processed CEL files can be found in the designated Archive folder.



# Opening the newly generated CHP file(s)

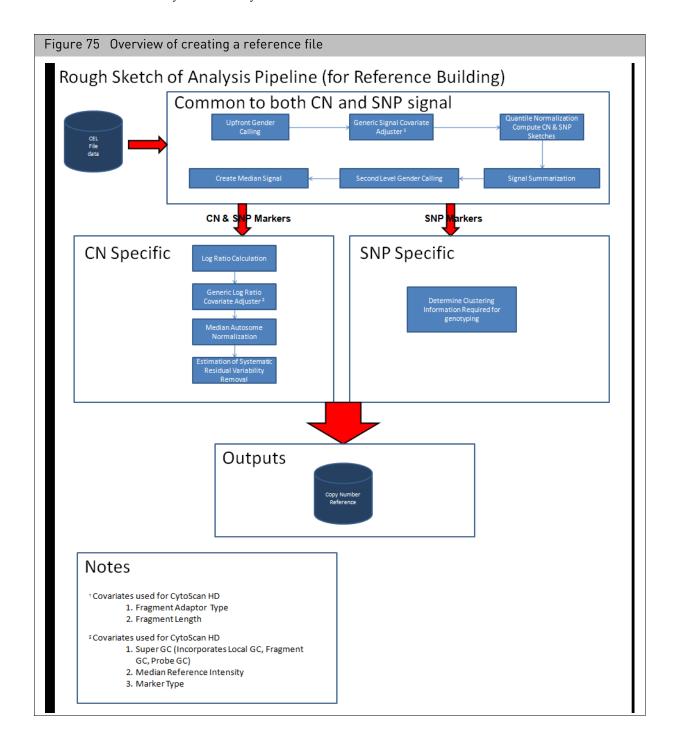
- From the ChAS Browser window, click File → Open.
   An Explorer window appears.
- 2. Navigate to your assigned Automatic CEL File Analysis output folder(s). See "Assigning your output folder(s)" on page 86.
- 3. Click or Ctrl click to highlight the CHP files you want to open in the ChAS Browser, then click **Open**.

The file(s) appear in the ChAS Browser.

# Reference files

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.

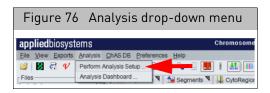
See Figure 75 for an overview of the analyses involved in creating a reference file for the CytoScan Arrays.



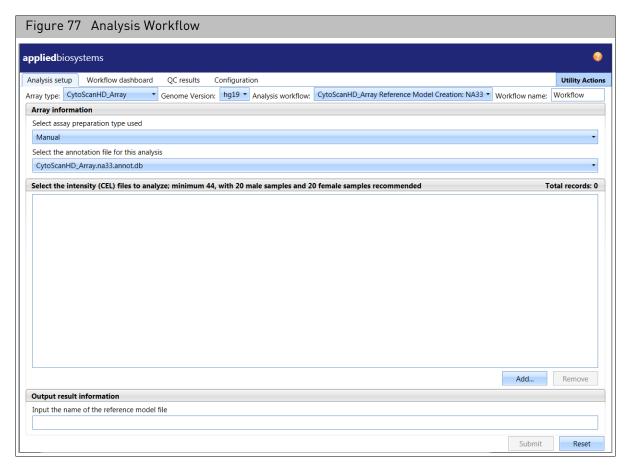
# Creating a reference file

**IMPORTANT!** When creating a CytoScan reference file, it is recommended that you use a minimum of 44 CEL files. These CEL files must include at least 20 male and 20 female samples.

1. From the Analysis menu, select **Perform Analysis Setup**. (Figure 76)



The Analysis Workflow window tab opens. (Figure 77)



2. From the **Select array type** drop-down list, click to select an array type (Example: CytoScanHD\_Array.

**Note:** The Select 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.

3. Select a Genome Build. (Example: hg19)

- 4. From the **Select analysis workflow** drop-down list, click to select an analysis workflow. (Example: CytoScanHD\_Array Reference Model Creation:NA33)
- 5. By default, the Set workflow name is **Workflow**. Click inside the Workflow's (upper right) text box to enter a different workflow name.
- Click the Assay Preparation type used drop-down button, then select one of the following:
  - Manual (CEL files whose assays were performed by hand)
  - Automation (CEL files whose assays were assisted by a robot)
  - Automation, Manual (a mix of CEL files whose assays were performed either by hand or with the assistance of a robot.

**Note:** The Annotation file for analysis is auto-selected once the array type, analysis workflow and assay preparation type used fields are selected.

- 7. From the **Select the annotation file for this analysis** drop-down list, verify the selection of the annotation file. (Example, CytoScanHD\_Array.na.33.annot.db)
- 8. At the Select the intensity (CEL) file(s) to analyze pane, click Add...

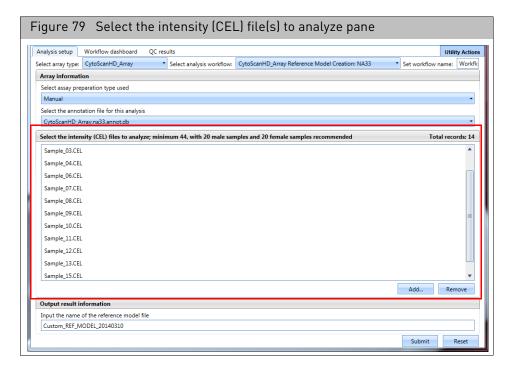
**IMPORTANT!** The same annotation file you used to create a Reference Model File MUST also be used with future Single Sample Analyses runs that utilize your created Reference Model File.

Figure 78 >CEL file folder Chromosome Analysis Suite Current directory C:\CytoScan\Data\Batch 1 Up one level Date Modified Туре X Sample\_01.CEL 8/8/2011 7:15 AM 66 MB AGCC CEL File 9e5653c3-de8... C:\Users\cgates ■ Sample\_02.CEL 10/14/2011 6:23 AM 66 MB AGCC CEL File 12f20ec2-0349.. \Desktop Sample\_03.CEL 7/31/2013 4:41 PM 66 MR AGCC CEL File 9e565f0a-9899 cb115160-647. X Sample\_04.CEL 5/17/2013 8:19 AM 66 MB AGCC CEL File X Sample\_06.CEL 5/28/2013 11:10 AM 66 MB AGCC CEL File 8efa3b98-481.. C:\Users\cgates X Sample\_07.CEL 5/28/2013 11:11 AM 66 MB AGCC CEL File 664b82e7-d43. ■ Sample\_08.CEL 11/23/2011 11:11 AM 66 MB AGCC CEL File de2455c4-a8a.. C:\Cytoscan Sample\_01.CEL;Sample\_02.CEL;Sample\_03.CEL;Sample\_04.CEL;Sample\_06.CEL;Sample Open Cancel Files of type: Intensity (CEL) file(s)

The following window appears: (Figure 78)

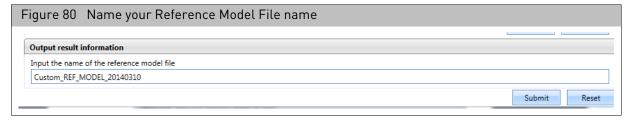
9. Single click, Ctrl click, or Shift click (to select multiple CEL files), then click Open.

The **Select the intensity (CEL) file(s) to analyze** pane is now populated with your CEL files. (Figure 79)



To remove a CEL file from this list, click to highlight it, then click Remove

**10**. At the **Output result information** pane, enter a name for your reference model file. (Figure 80)



11. Click Submit



# Analysis workflow exports and QC tools

### This chapter includes:

- "Displaying and exporting data from the analysis workflow"
- "Mendelian error checking" on page 113
- "Analysis workflow troubleshooting" on page 117

# Displaying and exporting data from the analysis workflow

# Adding files to the QC results table

- Click the QC results tab. (Figure 81)
   The QC results window tab appears.
- 2. Click Add Files.

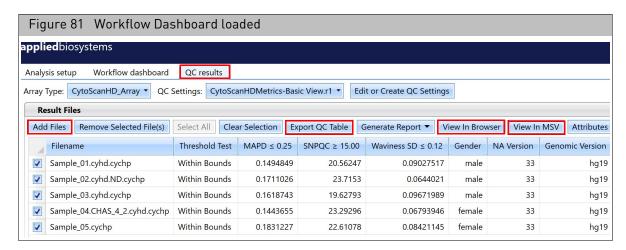
A Chromosome Analysis Suite window appears.

- 3. Navigate to your folder's location, then select the xxCHP files you want to add.
- 4. Click Open.

Your selected files appear in the Export QC Table window tab, as shown in Figure 81.

# Exporting QC table information

- 1. Check the adjacent check box next to the file(s) you want to export or click the **Select All** button (atop the check boxes) to auto-check all the displayed files.
- 2. Click on the **Export QC Table** tab.



An Explorer window appears.

- 3. Navigate to the export location you want, then enter a name for your QC file.
- 4. Click Save.

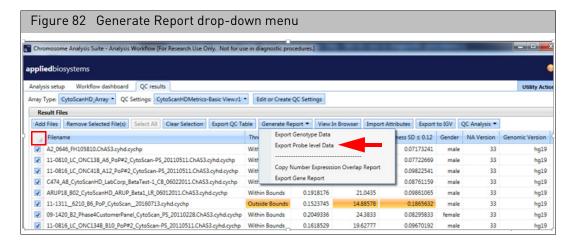
# Viewing analysis files in the ChAS browser

# Exporting probelevel data

- 1. Check the adjacent box next to the file(s) you want to open in the ChAS Browser.
- 2. Click the **View in Browser** button to open the files in the ChAS Browser or click the **View in MSV** button to open the files in the Multi-Sample Viewer. For more information on the MSV see, the RHAS User Guide.

# Exporting CytoScan Probe-Level Data

- 1. In the QC Results tab, click the check box adjacent to the Results File(s) you want to generate a report for or click the Select All button (atop the check boxes), as shown in Figure 82.
- 2. To export probe-level data, click the **Generate Report** drop-down. (Figure 82) The following export reporting options appear: (Figure 82)



3. Click to select **Export Probe Level Data**. (Figure 82)

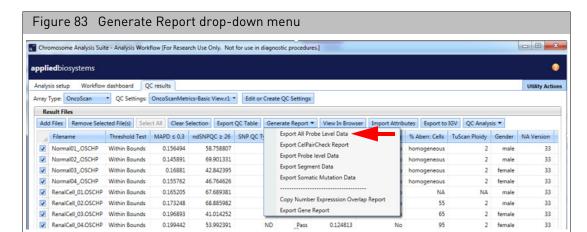
Your previously assigned Output folder file window appears.

**Note:** The default root filename is Result. Click inside the File Name field to enter a different root filename.

4. Enter a File Name for your text (tab-delimited) reporting file, then click **Save**, or navigate to different save location.

#### Exporting OncoScan probe-level data

- 1. In the **QC Results** tab, click each check box next to the Results File(s) you want to generate a report for. Click the **Select All** button (atop of the check boxes) (Figure 83) to auto-check all the displayed files.
- 2. To export probe-level data, click the **Generate Report** drop-down.



The following export reporting options appear: (Figure 83)

3. If you want to export all 4 available reports at once, click to select **Export All Probe Level Data**. (Figure 83) Otherwise, click to select the specific report(s) you want export.

The appropriate (previously assigned) folder file window appears.

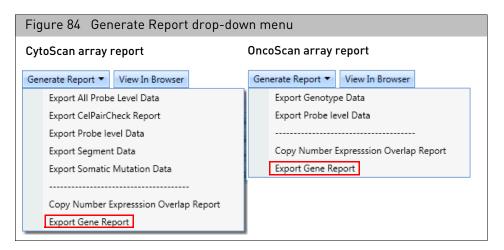
**Note:** The default root filename is Result. Click inside the File Name field to enter a different root filename.

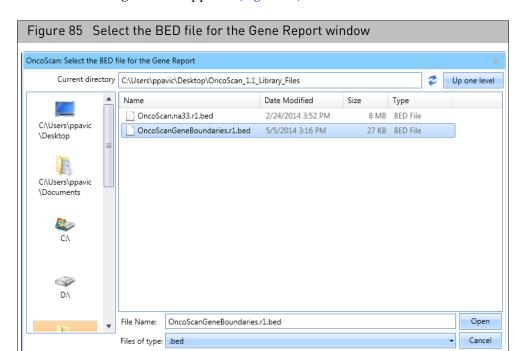
4. Enter a File Name for your reporting file, then click **Save** or navigate to a different save location.

Exporting a gene report (CytoScan or OncoScan arrays)

This report summarizes the copy number segments that overlap user defined regions of interest (e.g., Genes) as defined in the selected BED file.

 From the QC Results tab, click the Generate Report drop-down menu and select Export Gene Report. (Figure 84)



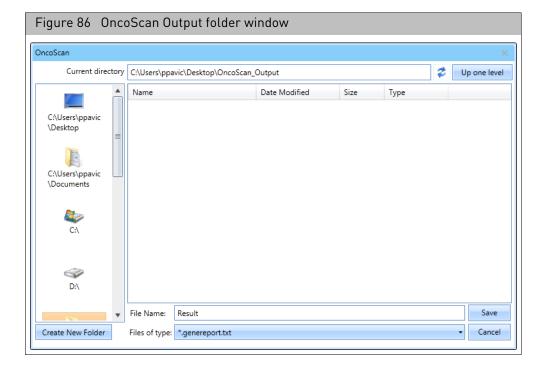


The following window appears. (Figure 85)

2. Click to select the appropriate BED file, then click **Open**.

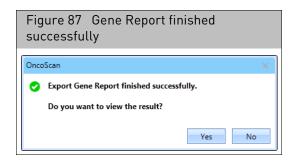
**Note:** Any BED file can be used to generate the Gene Report on any regions of interest contained within the BED file.

Your previously assigned Output folder file window appears. (Figure 86)



3. The default root filename is Result. Click inside the File Name field to enter a different root filename, then click **Save**.

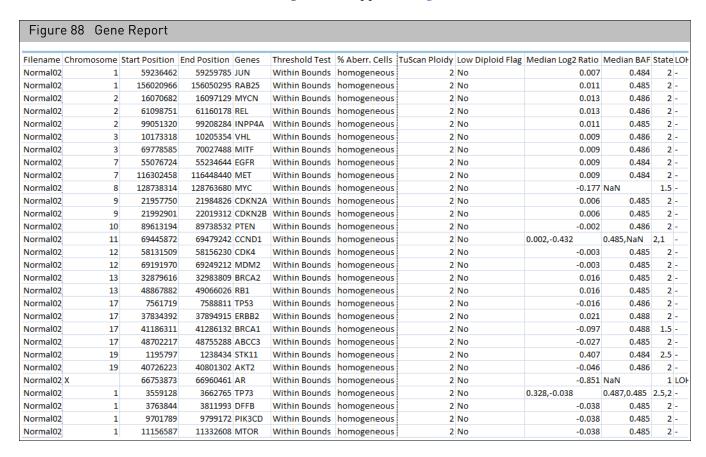
A progress bar appears while your report generates, followed by a message window. (Figure 87)



4. Click Yes.

The Results Output folder window appears.

5. Locate the Gene Report text file, then open it in Microsoft Excel. The following window appears. (Figure 88)



Gene Report Column	Description
Filename	Name of the xxCHP file containing the data.
Chromosome	Chromosome on which the probeset is located.
Start Position	Start position of gene or region as defined in the bed file.
End Position	End position of gene or region as defined in the bed file.
Genes	This column is populated from the name column of the bed file. In most cases, it will contain gene names.
Threshold Test	Displays Outside Bounds if any of the QC metrics fail to meet a threshold test. For more information on thresholds, see "To create your own custom QC setting:" on page 57.
% Aberr.Cells (OSCHP only)	If % AC = 100%, we return "homogeneous" because it could be 100% normal or 100% tumor. If % AC =NA, the percent aberrant cells could not be determined and TuScan returns non-integer CN calls. This metric is an algorithmically determined estimate of the % of aberrant cells in the sample.
TuScan Ploidy (OSCHP bnly)	TuScan Ploidy is the most likely ploidy state of the tumor before additional aberrations occurred. TuScan Ploidy is assigned the median CN state of all markers, provided that %AC could be determined and integer copy numbers are returned. If %AC cannot be determined, NA (Not Available) is reported for both ploidy and %AC.
Low Diploid Flag (OSCHP only)	An essential part of the algorithm is the identification of "normal diploid" markers in the cancer samples. This is particularly important in highly aberrated samples. The normal diploid markers are used to calibrate the signals so that "normal diploid markers" result in a log2 ratio of 0 (e.g. copy number 2). The algorithm might later determine that the "normal diploid" markers identified really correspond to (for example) CN=4. In this case the log2 ratio gets readjusted and TuScan ploidy will report 4. Occasionally (in about 2% of samples) the algorithm cannot identify a sufficient number of "normal diploid" markers and no "normal diploid calibration occurs. This event triggers "low diploid flag" = YES. In this case the user needs to carefully examine the log2 ratios and verify if re-centering is necessary.
Median Log2 Ratio (OSCHP only)	Log2 Ratio is the log2 ratio of the normalized intensity of the sample over the normalized intensity of a reference with further correction for sample specific variation. The Median Log2 Ratio is computed for each segment.
Median BAF (OSCHP only)	B-allele frequency (BAF) is (Signal (B)/{Signal(A) + Signal(B), where signal (A) is the signal from the AT chip and signal (B) is the signal from the G/C chip. Median BAF is reported for each segment and is the median BAF of the markers identified as heterozygous, after mirroring any marker BAFs above 0.5 to the equivalent value below 0.5. If the number of heterozygous markers in the segment is below 10 or the percent of homozygous markers is above 85% no value is reported,
State	This is a comma separated list of the copy number state of the segments that overlap the gene or region.
LOH	Flag to indicate whether the gene or region is in a Loss of Heterozygosity region (0=No, 1=Yes).

# Exporting a XON region report (CytoScan XON only)

This export provides XON Region gains/loses for a defined list of genes.

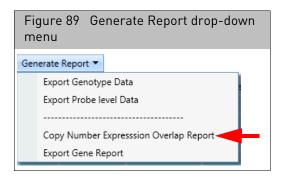
- 1. From the QC Results tab, select the check box for the sample file(s) for which you would like to generate an XON Region report.
- 2. Click on the Generate Report drop down menu and select Export XON Region Report.
- 3. In the browse window, click to select the appropriate aed/bed file, then click **Open**.
- 4. Click the browse button to assign an output path. Note that the default root filename is **Result**.
- 5. Optional: Click inside the Filename field to enter a difference root filename, then click **Save**.
- 6. Click the appropriate radio button (for each XNCHP file) to either export all data into a single text file or into individual text files.
- 7. Locate the XON Region Report text file, then open it in Microsoft Excel.

XON Region Report Column	Description
Filename	Name of the XNCHP file containing the data
Gene	This column is populated from the name column of the bed file. In most cases, it will contain gene names.
Chromosome	Chromosome on which the probeset is located.
Gene Start Position	Start position of the gene or region as defined in the bed file.
Gene End Position	End position of gene or region as defined in the bed file.
XON Region Start Position	Start position of the Exon Region segment as defined in the XNCHP file.
XON Region Stop Position	End position of the Exon Region segment as defined in the XNCHP file.
XON Region Type	Gain or Loss
Size (bp)	Size of the Exon Region Segment.
Marker Count	Number of markers in the Exon Region Segment.

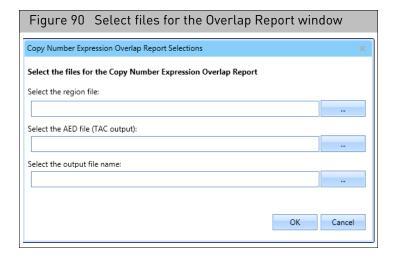
# Exporting a copy number expression overlap report

This report summarizes the copy number segments with the fold change from expression data that overlap user defined regions of interest (e.g. Genes) as defined in the selected AED or BED file.

1. In the QC Results tab, click on the Generate Report drop-down menu, then select Copy Number Expression Overlap Report. (Figure 89)



The following window appears: (Figure 90)



2. Click the Select the region file **Browse** button to navigate to and select the appropriate file.

**Note:** For the Regions file, you can use the default bed files provided in the in the library files for use with the Copy Number Expression Overlap report. You may also create your own AED/BED file containing your custom regions of interest.

- 3. Click the Select the AED file (TAC output) **Browse** button to navigate to and select the appropriate file. Refer to the Transcriptome Analysis Console (TAC) 4.0 User Manual for analyzing and exporting expression data as an AED file.
- 4. Click the Select output file name **Browse** button to navigate to an existing report location, or click inside the text field to enter a different root filename (other than the default Results filename), then click **OK**.
  - A progress bar appears while your report generates.
- 5. Locate the **cnexoverlapreport.txt** file, then double-click on it to open it. It is recommended to open the tab delimited file with Excel for easier viewing.

# The following window appears. (Figure 91)

Figure 91 Expression Gene Report													
GexCN9yearFFPELungReport20	141006 cnex	poverlaprep	ort - Notepad	-									
File Edit Format View Help													
Filename Gene Segment End Position BAF TCID Fold Ch 0009457_B04.OSCHP		Size (b	Gene Start Posi D) Percent art Position 59246462	Gene Ov	erlap d Positio	d Positio % Aberr. on 3324		TuScan I	Ploidy TCID Ge	ize (bp) Low Dip ne Symbo 1101959	loid Flag l	e 9 1094417	Seg LOH
0.004 0.486 0009457_B04.OSCHP	RAB25	708.hg.1 1	-2.26 59246460 156030966	0 1560402	59249999 95	9 9330	0	JUN 61	1102463	59	2076560	50	974
NO - 0.006 0009457_B04.0SCHP - 0.002 0.487	0.488 MYCN	2	16080682	1608712	9	6448	2	64	21494	6195371	8	6193222	5
0009457_B04.05CHP - 0.002 0.487	REL	2	61108751	6115017	8	41428	2	64	21494	6195371	8	6193222	5
0009457_B04.OSCHP NO0.001	INPP4A 0.487	2	99061320	9919828	-	136965	2	66	6235017		2430523		180
0009457_B04.05CHP LOH -0.387 0.316	VHL -	3	10183318	1019535	-	12037	1	67	63411	1978525		1977891	
0009457_B04.OSCHP LOH -0.387 0.316 0009457_B04.OSCHP	MITF TC030004 EGFR	3 4 <b>19.</b> hg.1 7	69788585 3.85 69788586 55086724	7001748 6 5522464	7001748	228904 8 137921	1 0 3	67 MITF 96	63411 1659291	1978525	64 1591184	1977891 13	.54 142
NO - 0.28 0009457_B04.05CHP	0.396 MET		328. hg. 1//TC07007 116312458		1	-2.36//-			1659291 1659291	300		3//55238	
NO - 0.28 0009457_B04.05CHP NO - 0.28	0.396 MYC 0.399	8 TC080000	 128748314 749.hg.1 2.32	- 1287536 1287476		- 5367 12875368	3 80	101 0	1135870 MYC	68	1462927	34	327
0009457_B04.OSCHP LOH -0.393 0.315	CDKN2A	9	21967750	2197482	-	7077	1	102		1410547		1408500	
0009457_B04.OSCHP LOH -0.393 0.315 0009457_B04.OSCHP	CDKN2B - PTEN	9 - 10	22002901 - 89623194	2200931 - 8972853	_	6412 105339	1	102 103	204738 126070	1410547		1408500 1353082	
LOH -0.396 0.316 0009457_B04.05CHP	CCND1	11	69455872	6946924	2	13371	2	104	192764	8892962		8873686	
0.009 0.485 0009457_B04.05CHP	CDK4	718.hg.1 12	3.15 69455855 58141509	5 5814623	69469242 0	2 4722	0 1	CCND1 109	3790298	8	1338181	15	95
NO LOH -0.388 0009457_B04.05CHP NO LOH -0.388	0.313 MDM2 0.313	12 TC12001	 69201970 704.hg.1 -2.17	6923921 6923506		37243 69237017	<b>1</b>	109 0.00012	3790298	8_	1338181	15	95
0009457_B04.OSCHP NO0.008	BRCA2 0.487	13	32889616	3297380	9	84194	2	110	1908482		1111353		920
0009457_B04.OSCHP NO0.008	RB1 0.487	13	48877882	4905602	6 - 2	178145	2 400959	110	1908482		1111353		920
0009457_B04.05CHP 0.244 TC17001094.hg.1 0009457_B04.05CHP LOH -0.006 0.244	TP53 //TC1700 ERBB2	17 2468.hg.: 17	7571719 7578811 L 2.21//2 37844392		7565097	124 //7571720 40524		80263427 7590868, 124	/ //759086: 400959	79862469 3 8026342	0//0	0.01 TP53//- 7986246	
0009457_B04.05CHP LOH -0.006 0.244	BRCA1	17 -	41196311	4127613	2_	79822	2	124	400959	8026342	7	7986246	9
0009457_B04.05CHP LOH -0.006 0.244	ABCC3	<b>1</b> 7	48712217	4874528	8	33072	2	124	400959	8026342	7	7986246	9
0009457_B04.OSCHP	STK11	19 -	1205797 1228434		2	126	247232		1765246		55	2	No
0009457_B04.OSCHP 0.021 0.487	AKT2	19	40736223	4079130	-	55080	2	128		5909323		5678996	
0009457_B04.OSCHP LOH -0.808 NAN 0009457_B04.OSCHP	AR - TP73	X - 1	66763873  3569128 3652765	6695046 - 83638	1 - 2	186589 59	754192	147 11019590		1549294		1521161 0.08	.26 55
0.486 0009457_B04.05CHP	DFFB	1	3773844 3801993		2	59		11019590		1094417		0.08	55
0.486	-	=					. 3,1232						

Expression Gene Report Column	Description
Filename	Name of the OSCHP file containing the data
Gene	This column is populated from the name column of the bed file. In most cases, it will contain gene names.
Chromosome	Chromosome on which the probeset is located.
Gene Start Position	Start position of gene or region as defined in the bed file.
Gene End Position	End position of gene or region as defined in the bed file.
CN State	This is a comma separated list of the copy number state of the segments that overlap the gene or region.

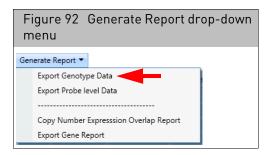
Expression Gene Report Column	Description
Segment ID	The unique identifier for the copy number segment.
Segment Start Position	Start position of the overlapping copy number segment in the xxCHP file.
Segment End Position	End position of the overlapping copy number segment in the xxCHP file.
Segment Size (bp)	The Segment Stop Position minus The Segment Start Position.
Percent Gene Overlap	Gene Boundary Size/Segment Size.
% Aberr.Cells (OncoScan only)	If % AC = 100%, we return "homogeneous" because it could be 100% normal or 100% tumor. If % AC = NA, the percent aberrant cells could not be determined and TuScan returns non-integer CN calls. This metric is an algorithmically determined estimate of the % of aberrant cells in the sample.
TuScan Ploidy (OncoScan only)	TuScan Ploidy is the most likely ploidy state of the tumor before additional aberrations occurred. Algorithmically it is the CN state of the markers identified by the algorithm as normal diploid before %AC and ploidy are determined. When a high ploidy is determined the "normal diploid" is deemed to correspond to a higher CN and the log2 ratio gets adjusted appropriately. If ploidy cannot be determined NA (Not Available) is reported.
Low Diploid Flag (OncoScan only)	An essential part of the algorithm is the identification of "normal diploid" markers in the cancer samples. This is particularly important in highly aberrated samples. The normal diploid markers are used to calibrate the signals so that "normal diploid markers" result in a log2 ratio of 0 (e.g. copy number 2). The algorithm might later determine that the "normal diploid" markers identified really correspond to (for example) CN=4. In this case the log2 ratio gets readjusted and TuScan ploidy will report 4. Occasionally (in about 2% of samples) the algorithm cannot identify a sufficient number of "normal diploid" markers and no "normal diploid calibration occurs. This event triggers "low diploid flag" = YES. In this case the user needs to carefully examine the log2 ratios and verify if re-centering is necessary.
Median Log2 Ratio (OncoScan only)	Log2 Ratio is the log2 ratio of the normalized intensity of the sample over the normalized intensity of a reference with further correction for sample specific variation. The Median Log2 Ratio is computed for each segment.
Median BAF (OncoScan only)	B-allele frequency (BAF) is (Signal (B)/{Signal(A) + Signal(B), where signal (A) is the signal from the AT chip and signal (B) is the signal from the G/C chip. Median BAF is reported for each segment and is the median BAF of the markers identified as heterozygous, after mirroring any marker BAFs above 0.5 to the equivalent value below 0.5. If the number of heterozygous markers in the segment is below 10 or the percent of homozygous markers is above 85% no value is reported,
LOH	Flag to indicate whether the gene or region is in a Loss of Heterozygosity region (0=No, 1=Yes).
Fold Change	The level of fold change as determined from the TAC software.
TCID	The Transcript Cluster ID overlapping the Gene or Region defined in the bed file.
TCID Start Position	Start Position: Start position of the overlapping TCID(s).
TCID End Position	End Position of the overlapping TCID(s).

Expression Gene Report Column	Description
TCID P-Value	ANOVA p-value (Condition1 vs Condition2)
TCID GeneSymbol	Gene Symbol: the Gene Symbol assigned to the TCIS(s) based on the TAC analysis.

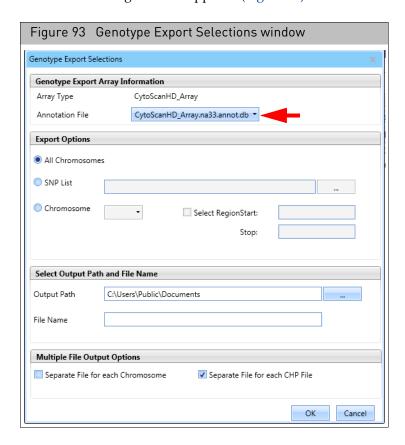
# Exporting genotype data using the analysis workflow

**Note:** For exporting genotypes from an entire sample or multiple samples at one time, exporting genotypes from the Analysis Workflow is generally faster than from the ChAS browser.

1. In the QC Results tab, click on the Generate Report drop-down menu, then select Export Genotyping Data. (Figure 92)



The following window appears: (Figure 93)



# **Exporting options**

**Note:** The default export is All Chromosomes.

## **Exporting a SNP List**

- Click the Annotation File drop-down menu (Figure 93), then click to select which array annotation file you want to use for exporting SNP information (along with the genotypes).
- 2. Click the **SNP List** radio button, then click the SNP List **Browse** button. The Select SNP List window appears. An example SNP List can be seen in Figure 386 on page 349.
- 3. Navigate to, then click to select the SNP List you want to export.
- 4. Click Open.

Your SNP List is now set for exporting.

## Exporting a specific chromosome

- 1. Click the **Chromosome** radio button.
- 2. Click the **Chromosome** drop-down menu, then click to select the specific number or chromosome type you want to export.
- 3. Optional: Click to check the **Select Region** check box, then enter a **Start** and **Stop** value in the provided text fields.

# Selecting an output path and filename

- 1. Click the Output Path **Browse** button.
  - The **Select Output Path for the Genotype Export Report** appears.
- 2. Navigate to an existing report location, or click inside the text field to enter a different root filename, then click **OK**.
- 3. Click inside the File Name text field to enter a report name.

### Multiple file output options

- 1. Click to check the **Separate File for each Chromosome** and/or **Separate File for each CHP File** check box.
- 2. Click OK.
  - A Please Wait...Exporting Genotype Data progress bar appears.
- 3. After a few moments, the **Export Genotype Data finished successfully** window appears. Click **Yes** to view the report, click **No** to view the report later.
  - Clicking **Yes** opens your previously assigned Output folder window. Locate the newly exported file, then double-click on it to open it. It is recommended to open the tab-delimited text file with Excel for easier viewing.

# Saving and importing attributes

Sample attributes can be added to the Results table for use in the PCA and Export IGV functions. A tab-delimited text file containing sample attributes can also be added to the Results table, however, column A must contain the name of the xxCHP file. Subsequent columns can contain other sample attributes. Alternatively, sample attributes listed in an ARR file generated in AGCC will also be automatically displayed in the Results table (as long as the ARR and xxCHP files are located in the same directory).

- 1. In the ChAS Analysis Workflow, select the **QC Results** tab.
- 2. Click Add Files to navigate to and load your CHP files.

The **Attributes** button is enabled.

- 3. Click on the **Attributes** drop down, select **Import**, then navigate to the corresponding TXT file.
- 4. Click Open.

The attributes now appear in the table.

5. Optional: To save an individual text file of sample attributes for each loaded chp file, click the **Attributes** drop-down, then select **Save**.

# Exporting to Integrative Genomics Viewer (IGV)

The ChAS Analysis Workflow enables you to export a variety of graphs for viewing in IGV. To access this viewer, go to: http://software.broadinstitute.org/software/igv/

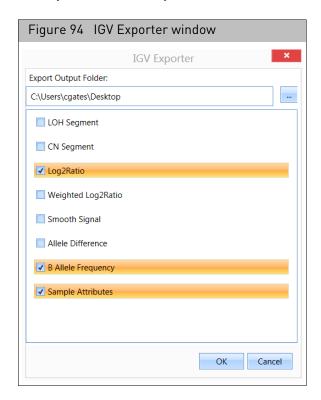
- 1. In the ChAS Analysis Workflow, click the **QC Results** tab.
- 2. Load results files by clicking on **Add Files**, then navigate to and highlight the CHP files.
- 3. Click Open.
- Select files by either clicking the Select All, or checking each filename's adjacent check box.
- 5. Click Export to IGV.

The IGV Exporter window appears. (Figure 94)

- 6. Click the Browse button to assign an output folder.
- 7. From the IGV Exporter window, click the check box(es) adjacent to the data you want to export, as shown in Figure 94.
- 8. Click OK.

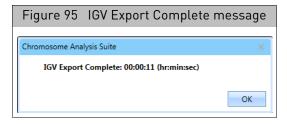
**Note:** To include sample attribute information in the IGV Export, click on **Attribute**  $\rightarrow$  **Save** prior to running the IGV Export.

**Note:** XON Region segments are exported as CN segments when working with CytoScan XON arrays.



**Note:** The export process may take several minutes to complete, as it is dependent on your sample(s) and data type(s).

If the export was successful, an IGV Export Complete message appears. (Figure 95) Acknowledge the message, then click **OK**.



Your data is now ready to be imported into IGV's browser.

For more information on importing and viewing your data in IGV, go to:

http://software.broadinstitute.org/software/igv/UserGuide

# Principle component analysis

ChAS Analysis Workflow enables you to perform Principle Component Analysis (PCA) on signal (CHP) data. PCA identifies a new set of variables (PCA1, PCA2, and PCA3) that account for a majority of the variance in the original data set. The first principal component (PCA1) captures as much variability in the data as possible. PCA2 captures as much of the remaining variability (not accounted for by PCA1) as possible. PCA3 captures as much of the remaining variability (not accounted for by PCA2) as possible.

# PCA plot generation

- 1. In the ChAS Analysis Workflow, select the QC Results tab. Load results files by clicking on Add Files, and navigating to xxCHP files, then click **Open**.
  - When files are loaded, the Import Attributes, Export to IGV and QC Analysis buttons will be enabled.
- 2. Select files, by either clicking **Select All** button or selecting files individually by ticking box to left of file.
- 3. Click on the QC Analysis button and in the resulting drop down menu, select PCA.

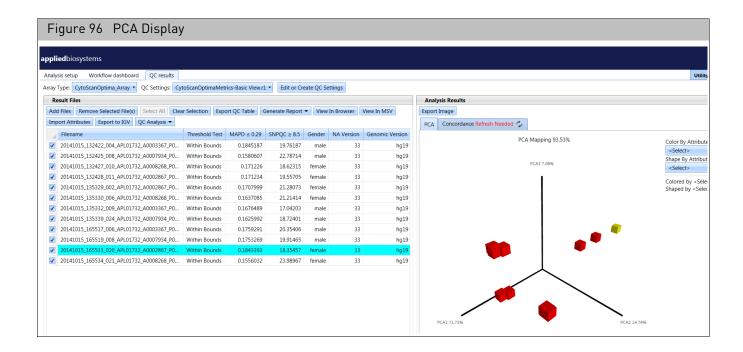
The 3-dimensional PCA graph(s) appears. The graph axes represent the top three variables (PCA1, PCA2, and PCA3) that account for the majority of the variability among the samples.

## Sample display options

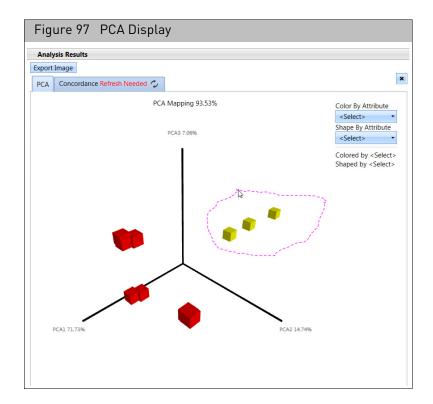
By default, the table is set to a vertical view. Click to change the view to a horizontal view.

1. The file can be identified by clicking on an icon in the plot.

The corresponding file in the table is highlighted, as shown in Figure 96.



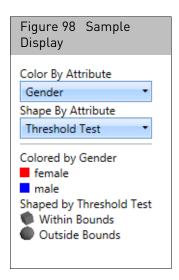
2. Icons can be lassoed by left-clicking and drawing circle around sample(s) of interest, as shown in Figure 97. This action also highlights the corresponding files in the table.



3. Points on the plot can also be hidden by right-clicking, then selecting the appropriate option.

**Note:** The selected samples are only hidden from the plot.

4. Use the drop-down menus (Figure 98) to select attributes for display by Color By Attribute and by Shape By Attribute.



#### Additional PCA graph display options

- To rotate the graph, right-click, then drag the graph to change its view perspective.
- Click **Export Image** to save the displayed plot as a PNG graphics file.

# Concordance checks

The ChAS Analysis Workflow enables you to perform pairwise comparison concordance checks on genotype calls for all selected samples.

The concordance between all pairwise comparisons for the samples in the results table are reported.

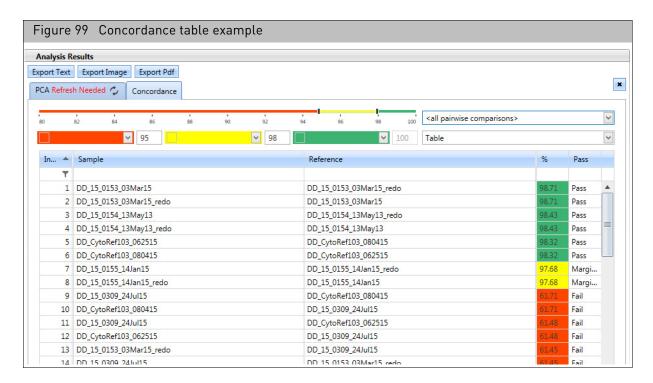
A reference sample can be selected. Once selected, concordances are displayed.

Compare to reference allows you to compare every sample to a single reference file.

#### Performing a concordance check

- 1. In the ChAS Analysis Workflow, select the **QC Results** tab.
- 2. Load results files by clicking on **Add Files**, then navigate to xxCHP files. After the files are loaded Import Attributes, Export to IGV, and QC Analysis buttons are enabled.
- 3. Select files, by either clicking **Select All** button or by clicking each file's (left) check box.
- 4. Click on the **QC Analysis** button. A drop-down menu appears.
- 5. Click Concordance.

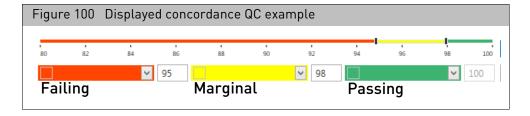
The concordance table with sample, reference and percent concordance is generated, as shown in Figure 99.



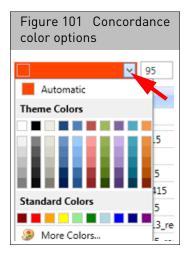
#### Concordance table filter and display options

- To compare to reference sample, select the sample from drop down-menu.

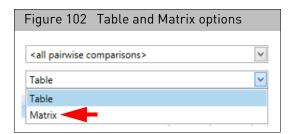
  | <all pairwise comparisons> | V |
- The default values are set at >98%, 95-98%, and below 95%.
- The three categories of QC are passing, failing, and marginal.
- Each QC category is represented by its own unique color, as shown in Figure 100.



• Colors in the table can be changed by clicking on the color drop-down (Figure 101), then choosing a different color from the pallet.

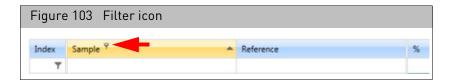


- The default values can be changed by either moving the hash marks on the line bar, or typing in a number in the designated category (Figure 101).
- The data can also be viewed in matrix format by selecting **Matrix** in the drop-down menu, as shown in Figure 102.

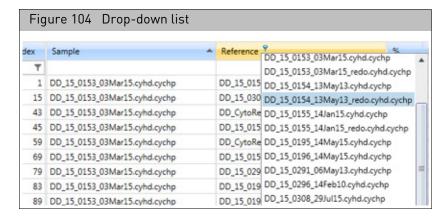


#### **Filtering**

At the table, place your mouse cursor over a Sample or Reference cell.
 A filter icon appears. (Figure 103)



Right-click on the Filter icon.A drop-down list appears. (Figure 104)



3. Click to select the samples you want to filter from the drop-down list.

#### Changing the view and/or order of sample and reference columns

Left-click the filter icon.
 A menu appears. (Figure 105)



2. Use the menu selections to customize your sample and reference columns.

#### Exporting the currently displayed table

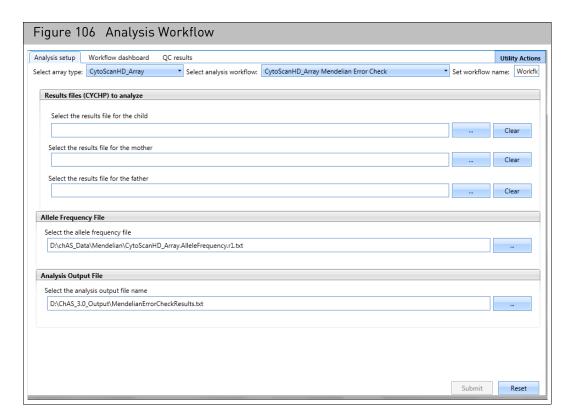
1. Click the appropriate, Export Text Export Image Export Pdf button, then export the file as you normally would.

### Mendelian error checking

**Note:** The Mendelian Error Checking feature is for the CytoScan family of arrays only.

# Running an error checking analysis

1. From the Analysis menu, select **Perform Analysis Setup.**The Analysis Workflow window opens. (Figure 106)



2. From the **Select array type** drop-down list, click to select an array type (Example: **CytoScanHD\_Array**.

**Note:** The **Select 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.

- From the Select analysis workflow drop down, click to select the appropriate Mendelian Error Check Workflow. (Example: CytoScanHD\_Array Mendelian Error Check)
- 4. By default, the Set workflow name is **Workflow**. Click inside the Workflow's (upper right) text box to enter a different workflow name.

Click each of the three Results files (CYCHP or XNCHP) to analyze Browse buttons to navigate to and select the appropriate Results File for Child, Mother, and Father.

You can also run a Mendelian Error Check using two Results files: **Child** and **Mother** or **Child** and **Father**.

6. Click the Allele Frequency File **Browse** button to navigate to and select the appropriate Allele text file.

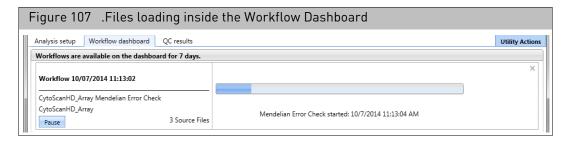
**Note:** An Allele Frequency file is provided in the library file package or you can create your own custom Allele Frequency File for use in this analysis.

 Click the Analysis Output File Browse button to navigate to and enter a name (tab-delimited text filename) for your Mendelian Error Check output result. By default, the Output name is based on the CYCHP/XNCHP filename used for the Child.

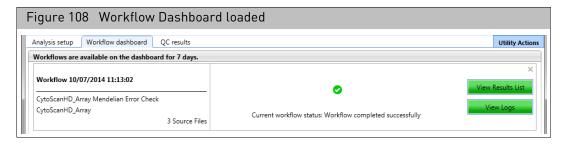
#### 8. Click Submit.

An error appears if a male is assigned as mother or a female is assigned as father or if the gender in the actual cychp file is wrong.

The Workflow Dashboard window appears. (Figure 107).



After loading is complete, a Workflow completed successfully message appears. (Figure 108)



#### 9. Click View Results List.

The Output folder window (you assigned earlier) appears.

- 10. Locate the newly created \*.cyhd.txt/cyex.txt (or \*.cychp.ND.txt) file, then double-click it to open it in MS Notepad to view it as tab-delimited text file, as you normally would.
- 11. Optional: If you want to view the Workflow's Log File, click **View Logs**. The C:\ProgramData\Affymetrix\ChAS\Log folder window appears.

Locate the newly created Log.txt file, then double-click it to open it in MS Notepad to view it as tab-delimited text file, as you normally would.

# Interpreting an error checking analysis

The Mendelian Error Check analysis provides two key points of information:

- 1. Are the input samples related?
  - Mother-Child
  - Father-Child

If the samples are related, the **Role Validity** equals 1. If the samples are not related, **Role Validity** equals 0. (Figure 109) The output also indicates which CYCHP/XNCHP file is assigned as the Mother, Father and Child (Index). The analysis also can be run as a DUO analysis. (Mother-Child or Father-Child).

1	#%GroupName=F	amilial			
_					
2	#%SetName=San	ipies			
3	#%Columns=4				
4	#%Rows=3				
5	SampleKey	CHPFilename	CHPID	Role	
6	0	A03_CytoScanHD_	V00001316-386a-4744	index	
7	1	A01_CytoScanHD_	00006534-6a66-4722	mother	
8	2	A02_CytoScanHD_	00006996-3b6b-4e53	father	
9	#%SetName=Rel	atednessTest			
10	#%Columns=5				
11	#%Rows=3				
12	AnalysisType	ReferenceSample	K FamilialSampleKey	RoleValidity	RoleIndexScore
13	0	(	2	1	80375.99
14	1	(	1	1	65454.53
15	2	(	2	1	48592.95

#### 2. Do any chromosomes have an elevated occurrence of Mendelian Errors?

In Figure 110, chromosome 15 has a higher error rate (7.2%) compared to the rest of the chromosomes in this trio. In this example, the mother has 0% errors on chromosome 15, whereas the father has 7%, indicating that both chromosome 15 (or some portion of) alleles were inherited from the mother. It is recommended to compare the genotypes on chromosome 15 for these samples using the Graphs tab. For more information, see "Graphs table" on page 345.

19	Chromosome	Display	MarkerCount	MIE-Trio	MIE-Mat	MIE-Pat	Percent-T	Percent-N	Percent-Pat
20	1	1	56482	119	21	38	0.21	0.04	0.07
21	2	2	62283	134	32	35	0.22	0.05	0.06
22	3	3	52106	108	14	39	0.21	0.03	0.07
23	4	4	49517	101	37	25	0.2	0.07	0.05
24	5	5	46205	108	25	36	0.23	0.05	0.08
25	6	6	51944	128	29	32	0.25	0.06	0.06
26	7	7	46413	131	39	31	0.28	0.08	0.07
27	8	8	38796	84	20	29	0.22	0.05	0.07
28	9	9	30622	72	24	20	0.24	0.08	0.07
29	10	10	35472	97	23	35	0.27	0.06	0.1
30	11	11	38846	105	37	28	0.27	0.1	0.07
31	12	12	33424	65	12	20	0.19	0.04	0.06
32	13	13	27733	45	12	17	0.16	0.04	0.06
33	14	14	26983	75	20	32	0.28	0.07	0.12
34	15	15	24981	1799	1	1752	7.2	0	7.01
35	16	16	20915	48	14	19	0.23	0.07	0.09
36	17	17	17465	44	13	10	0.25	0.07	0.06

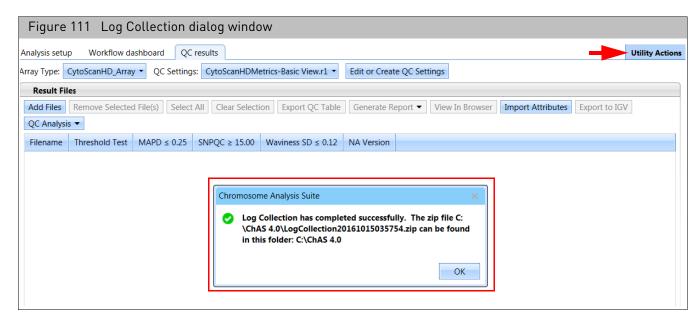
Mendelian Errors Report Column	Description			
Analysis Type	Provides the samples being run through the analysis based on the sample key.  0 = Proband, 1 = Mother, 2 = Father.			
Familial Sample Key	Tells you the parent for which relatedness is being tested.  1= Mother only used in the analysis.  2= Father used in the analysis (may be father only (duo) or trio).			
Role Validity	A logical value with 0 being False and 1 being True. If the Role Index Score is > 1000 then the Role Validity = 1 (likely related). If the Role Index Score is < 1000 then Role Validity = 0 (likely unrelated).			
Role Index Score	Role Index Score: This score is basically a Log Odds score that computes the probabilities of the observed genotype calls for the trio while accounting for potential genotyping error. Assume the following hypothesis: H1 – alleged father is true H2 – alleged father is random male. For each marker, compute a likelihood ratio of H1 vs H2. Sum all markers. In theory a value of zero means equally likely probability for either hypothesis. Positive means more likely Paternity related. Negative means more likely unrelated.			
Chromosome/ Display	Chromosome number			
Marker Count	Number of genotypeable SNPs on the chromosome.			
MIE Trio	Mendelian Inheritance Error for the Trio. (Ex 119 errors of 56482 SNPs).			
MIE Mat	Number of errors for Mom.			
MIE Pat	Number of errors for Dad.			
Percentage MIE Trio	Number of raw error turned into a percent 119/56482 *100.			
Percentage MIE Mat	Percent of errors for Mom.			
Percentage MIE Pat	Percent of errors for Dad.			

### Analysis workflow troubleshooting

Do the following if you experienced any issues or failures with your analysis:

- 1. Click the **Utility Actions** button (top right of the Analysis Workflow window)
- 2. Click to select **Log Collection**.

The following dialog window appears. (Figure 111)



- 3. Make note of the assigned zip folder filename and its location.
- Use Windows Explorer to navigate to the location. Example: C:\ProgramData\Affymetrix\ChAS\log
- 5. Locate the zip folder you noted earlier, then double-click on it to open it. The folder opens.

#### Log rollover

When the software determines that the log file for the Analysis Workflow (C:\ProgramData\Affymetrix\ChAS\log\AnalysisWorkflow.log) has reached a defined size (approximately 4MB), the following steps will be completed:

- A sub-folder will be created in C:\ProgramData\Affymetrix\ChAS\log called 'Log\*' (the '\*' denotes the current date and time).
- A zip file called **RolledLogFile\*.zip** is created in that folder. The '\*' is the same date and time used for the folder name. The files in the **C:\ProgramData\Affymetrix\ChAS\log** folder and all files found in the currently selected QC History Log folder will be included in this zip file.
- The Analysis Workflow files that are associated with analysis workflows that are no longer active on the Dashboard will be deleted from C:\ProgramData\Affymetrix\ChAS\log.
- A new AnalysisWorkflow.log file will be created in C:\ProgramData\Affymetrix\ChAS\log.

#### Log collection

When the Log Collection option is selected from the **Utility Actions** menu, a file called **LogCollection\*.zip** (the \* denotes the current date and time) is created in the folder C:\ProgramData\Affymetrix\ChAS\log.

This created file contains the full contents of the folder C:\ProgramData\Affymetrix\ChAS\log, including the log file for the browser (ChAS\_RUO.log).

If available, the sub-folders of C:\ProgramData\Affymetrix\ChAS\log and all files found in the currently selected QC History Log folder.

**Note:** All log files for the ChAS database, ChAS Browser and Analysis Workflow can be found in: \ProgramData\Affymetrix\ChAS\Log



# **Loading Data**

#### This chapter includes:

- "Introduction to loading data"
- "Loading files" on page 122
- "Copy number segment smoothing and joining (optional)" on page 128
- "XON segment merging" on page 133
- "Setting QC parameters in the ChAS browser" on page 134
- "Histogram data" on page 140

#### ChAS can display data from:

- CytoScan Array CYCHP/XNCHP files, generated in ChAS.
- CytoScan HTCMA files, generated in the RHAS.
- Genome-Wide Human SNP Array 6.0 CNCHP files, generated in Genotyping Console (GTC).
- OncoScan OSCHP files, generated in OncoScan Console or ChAS.
- ReproSeq Aneuploidy results (.zip) from Ion Reporter.

**Note:** When referring to steps that apply to both CytoScan CYCHP and SNP 6.0 CNCHP data files, the CHP files are described as CxCHP files. When referring to steps that apply to CytoScan CYCHP, CytoScan XNCHP, CytoScan HTCMA RHCHP, SNP 6.0 CNCHP data, and OncoScan files, the resultant files are described as xxCHP files.

### Introduction to loading data

The same steps are used to load results xxCHP files from CytoScan arrays, Genome-Wide Human SNP 6.0 Array, or OncoScan Arrays.

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

- 1. Loads 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 for CytoScan 750K and CytoScan HD arrays. Smoothing and joining are disabled for CytoScan Optima arrays. Smoothing and joining are OFF by default for OncoScan and CytoScan HTCMA arrays. The smoothing and joining settings are specific for each array type (for more details on smoothing and joining, see "Copy number segment smoothing and joining (optional)" on page 128).

- 3. Displays the segments and graph data:
  - Segment Data
    - Copy Number Gain/Loss
    - Mosaicism Copy Number Gain and Loss
    - Loss of Heterozygosity (LOH)
  - · Graph Data
    - Copy Number State
    - Log2 Ratio
    - Weighted Log2 Ratio
    - LOH
    - · Smooth Signal
    - Allele Difference
    - B-allele Frequency
    - Genotype Calls

When loading XNCHP files into ChAS for viewing from CytoScan XON arrays, the software:

- 1. Selects the segments in the XNCHP file to display as segments.
- 2. Displays the segments and graph data:
  - Segment Data
    - Loss of Heterozygosity
    - Exon Region Gain/Loss
  - Graph Data
    - Log2 Ratio
    - Weighted Log 2 Ratio
    - Smooth Signal
    - Loss of Heterozygosity (LOH)
    - Allele Difference
    - B-allele Frequency (BAF)
    - Genotype Calls

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

- 1. Selects the segments in the RHCHP file to displays as segments.
- 2. Displays the segments and graph data:
- Segments Data
  - Copy Number Gain/Loss
  - Loss of Heterozygosity
- Graph Data
  - Copy Number State
  - Log 2 Ratio
  - Smooth Signal
  - LOH
  - Allele Difference
  - B-allele Frequency

When loading SNP6 CNCHP files into ChAS for viewing, the software does the following:

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.

2. Applies any smoothing, removing, 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.

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

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

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. For more information see "Copy number segments on the X and Y chromosomes" on page 48.

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

When loading OncoScan OSCHP files for viewing, the software does the following:

1. Displays segments in the OSCHP created by the TuScan Copy number algorithm. For details on this algorithm, please refer to the **OncoScan Console User Guide** (P/N 703195) or Appendix G.

#### **IMPORTANT!** Smoothing and Joining are OFF by default for OSCHP files.

- 2. Displays the segments and graph data:
  - Segment Data
    - Copy Number Gain/Loss
    - Loss of Heterozygosity

- Graph Data
  - Copy Number State
  - Log2 Ratio
  - Weighted Log2 Ratio
  - Allele Difference
  - B-allele Frequency
  - LOH
  - · Smooth Signal
  - Variant/Somatic Mutation
  - Somatic Mutation (OncoScan FFPE Assay only)

When loading ReproSeq Aneuploidy data for viewing, the software does the following:

- 1. Displays segments from the Ion Reporter software. For details, refer to the Ion Reporter User Guide: https://ionreporter.thermofisher.com/ir/
- 2. Displays the whole genome sequencing tiles on the Copy Number State graph.

### Loading files

Loading xxCHP 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 for CxCHP files.

**IMPORTANT!** Smoothing and Joining are ON by default for CytoScan 750K and HD arrays. Both Smoothing and Joining are OFF by default for OncoScan, CytoScan HTCMA, and GenomeWide Human SNP 6.0 arrays. and is disabled for CytoScan Optima arrays and ReproSeq Aneuploidy data.

XON Segment Merging is turned ON by default for CytoScan XON arrays. For details, see "XON segment merging" on page 133.

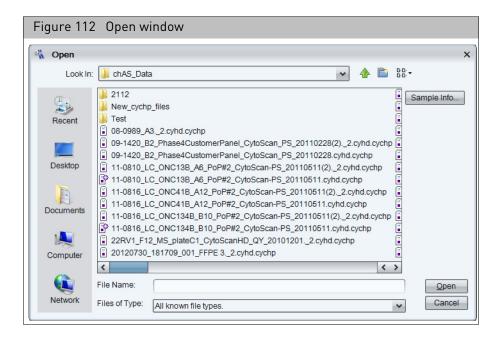
- 2. Select analysis results from:
  - CytoScan Arrays (generated by analyzing CEL files in ChAS)
  - CytoScan HTCMA arrays (generated from CEL files in the RHAS.
  - Genome-Wide Human SNP 6.0 Arrays (generated by analyzing CEL files in
  - OncoScan Arrays (generated by analyzing in ChAS)
  - ReproSeq Aneuploidy .zip files (generated in Ion Reporter)

You can also select region information files in AED and BED format for loading. Use the Open window (click the 📴 button) to load xxCHP 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 "Copy number segment smoothing and joining (optional)" on page 128 for more information.

3. Select File  $\rightarrow$  Open on the menu bar. Alternatively, click the File Open button  $\square$ .

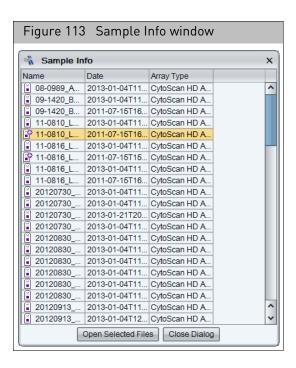


The Open window appears. (Figure 112)



 To view information about results, select one or more files, then click Sample Info.

The **Sample Info** window opens. (Figure 113)

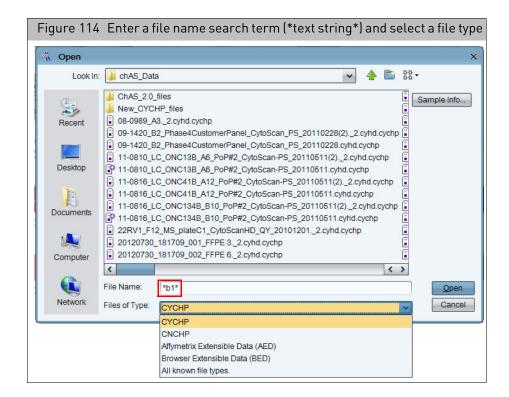


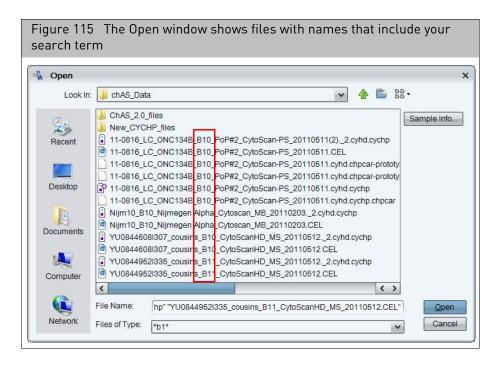
**Note:** If the xxCHP and ARR files are located in the same folder, the Sample Info window shows information about both the sample and the results. To load files from the **Sample Info** window, select the files, then click **Open Selected Files**.

### Using the search feature

- Click Close Dialog (Figure 113) to close the Sample Info window.
   The Open window appears.
- 2. Navigate to the folder with the files that you want to search for and load.
- 3. Enter a text string with an asterisk (\*) **before** and **after** the search term in the **File Name** field. See example in Figure 114 on page 124.
- 4. Select a file type from the drop-down list. See example in Figure 114 on page 124.
- 5. Click Open.

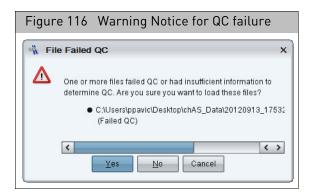
Files with names that include the search term are displayed in the **Open** window. See example in Figure 115 on page 125.





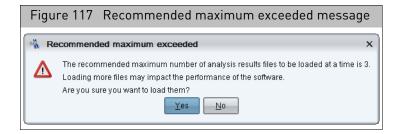
- 6. Select the files (you can use Shift click or CTRL click to select multiple files)
- 7. Click Open.

If any of the files fail the QC checks, a warning notice appears (Figure 116). You can click **Yes** to continue to load the files.



It is recommended you load no more than 3 files at one time.

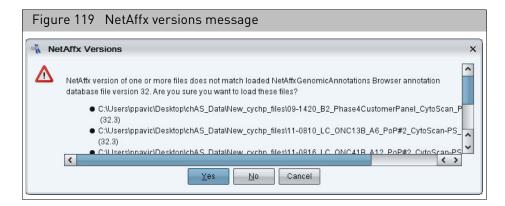
If the following warning message appears (Figure 117), click **OK** to acknowledge it or click **Cancel** to reselect a maximum of 3 files.



If the following warning message appears (Figure 118), acknowledge it, then click **OK**.

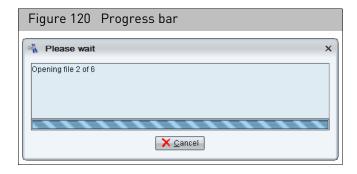


If the following warning message appears (Figure 119), click **Yes** to acknowledge it.

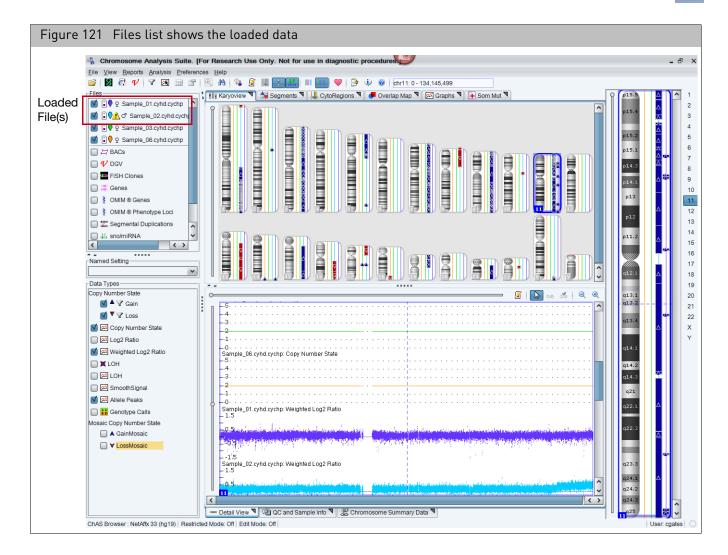


**Note:** The ChAS Browser allows loading different NetAffx versions at the same time (as long as the versions are all from all the same reference and genome builds). If NetAffx versions are from different builds of the genome (for example Hg18 and Hg19), The ChAS Browser does not load the files.

A progress bar appears (Figure 120)



After a few moments, the ChAS browser featuring your selected samples appears. The loaded files appear in the Files list pane. (Figure 121).



### Copy number segment smoothing and 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 for CytoScan 750k and HD Arrays.

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.

**IMPORTANT!** Smoothing and Joining affect only data loaded from CNCHP and CYCHP files. This ONLY applies to copy number data, NOT LOH or Mosaic types. Smoothing and Joining is OFF by default for CytoScan Optima, OncoScan, and ReproSeq Aneuploidy files.

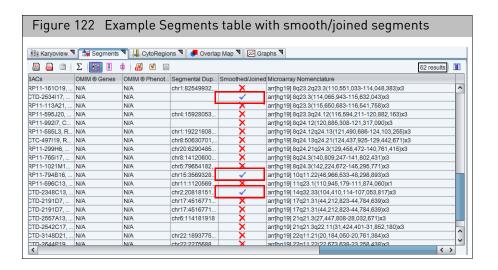
Segments which have been smoothed and/or joined are indicated by a blue check mark in the Smoothed/Joined column of the Segments table (Figure 122). The segment ID name indicates whether smoothing and/or joining has occurred. A red "X" indicates no smoothing or joining has been applied.

Viewing what segments were smoothed/joined

- 1. Click the Segments tab.
- 2. Use the horizontal scroll bar to move the window to the far right. The Smoothed/Joined column appears. (Figure 122)

#### For more information:

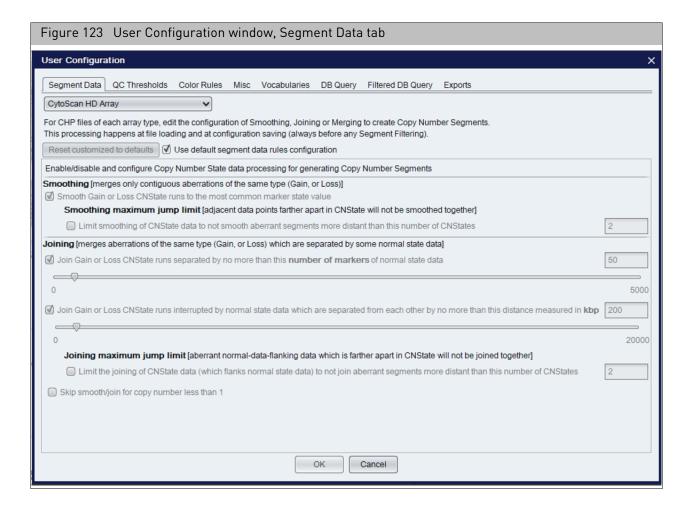
- See "About smoothing" on page 131.
- See "About joining" on page 132.



# Setting smoothing and joining parameters

- Click Preferences → Edit User Configuration.
   The User Configuration window appears.
- 2. Click the **Segment Data** tab.
- 3. Click the **Choose Array Type** drop-down menu to select the array type to view or edit its Smoothing and Joining settings. (Figure 123 on page 129)

**Note:** If you change the smoothing or joining parameters, the new rules are applied to the original, raw segments of CxCHP files which have not had modifications, Calls, Interpretations, or Inheritance made. CxCHP files which have had segments modified or had Calls, Interpretations or Inheritance made will have had their smoothing and joining parameters fixed and will not adopt changes made to the array type's Segment Data smoothing and joining settings.



### Segment data tab options and descriptions

Option	Description
Use default segment data rules configuration	For the CytoScan 750K and HD Arrays, the default smoothing and joining rules are:  - 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  - Skip segments from the smoothing operation for all arrays that have a CN < 1.  For SNP 6 arrays, 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	For the CytoScan 750K and HD Arrays, the default smoothing and joining rules are:  - 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 SNP 6 arrays, the default smoothing rule:  - Smooth Gain or Loss CNState runs to the most common marker value, then generate segments.
Limit the joining of CNState data (which flanks normal state data) to not join aberrant segments more distant than this number of CNStates	Smoothing to the most common marker state value is only applied to contiguous CNState runs of the same type (gain or loss).

**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 get recalculated. For more information:

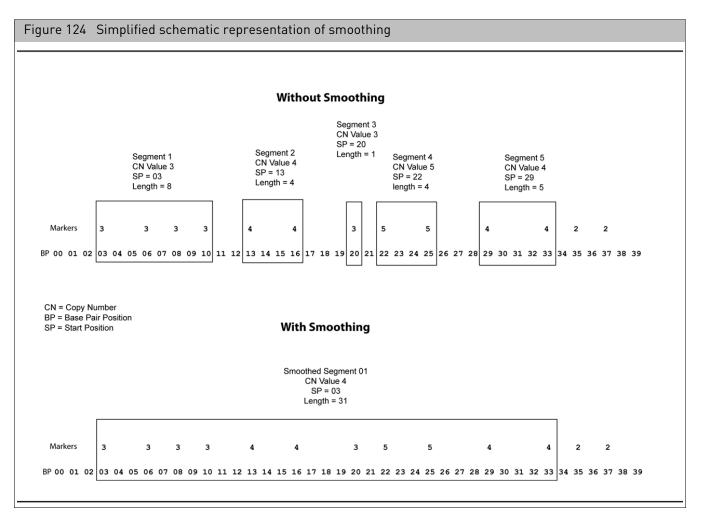
- See "About smoothing" on page 131.
- See "About joining" on page 132.

#### About smoothing

**Note:** The examples shown below 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 124)



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

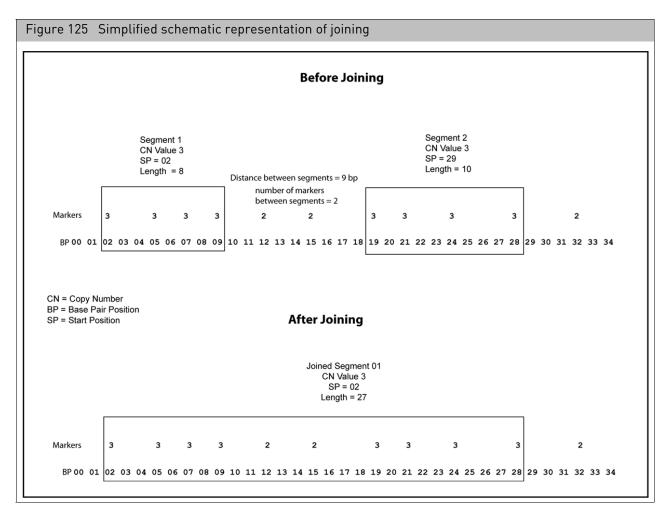
The median CNState of all the markers in the Segment is assigned as the Copy Number State value for the new smoothed segment. The median will thus always be either an integer or a half integer (like 3.5).

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).

#### About joining

The joining options enable 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 125).



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.

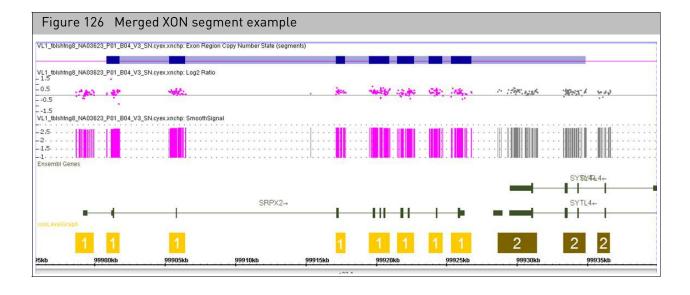
### XON segment merging

XON Merging is on by default. XON Merging combines consecutive XON segment calls of the same Type (XON Region Gains or XON Region Loss) into a larger segment. Once an XON segment of a different Type (normal state or gain/loss) is encountered, the XON Merge is terminated. A Merged XON segment is represented as the larger segment in the Segments Table.

The Merged XON segment is represented by the transparent rectangle. (Figure 126)

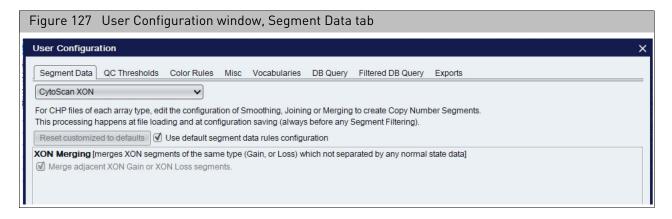
**Note:** The individual XON segments can still be viewed within the rectangle, provided the appropriate Levels are selected in the Filter Settings. In addition, the breakpoints of the Merged XON Segment will stop with an XON segment and may not line up exactly with the probe level data.

**Note:** A merged XON segment is assigned to the lowest XON annotation number within that merged segment. For example, if a merged XON segment overlaps XON Regions annotations of Level 1 and Level 3, the whole merged XON segment will be considered Level 1.



# Turning off XON merging

To turn off XON Merging, go to Preferences > User Configuration.
 The User Configuration window appears. (Figure 127)



- 6
- Click the Segment Data tab, then select CytoScan XON from the array dropdown.
- 3. Uncheck the **Use default segment data rules configuration** check box
- 4. Uncheck the **XON Merging** check box.
- 5. Close the ChAS browser, then reopen it. XON Merging is now off/disabled.

### Setting QC parameters in the ChAS browser

ChAS checks the analysis results files for certain QC values. The software notifies you if the QC parameters do not meet the thresholds.

**Note:** Custom QC metrics can be viewed in both the Analysis Workflow and the Browser. However, any custom settings you wish to use, have to be entered separately in both the Analysis Workflow and the Browser. To create custom QC settings in the Browser see below. To create custom QC settings in the Analysis Workflow, see "To create your own custom QC setting:" on page 57.

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

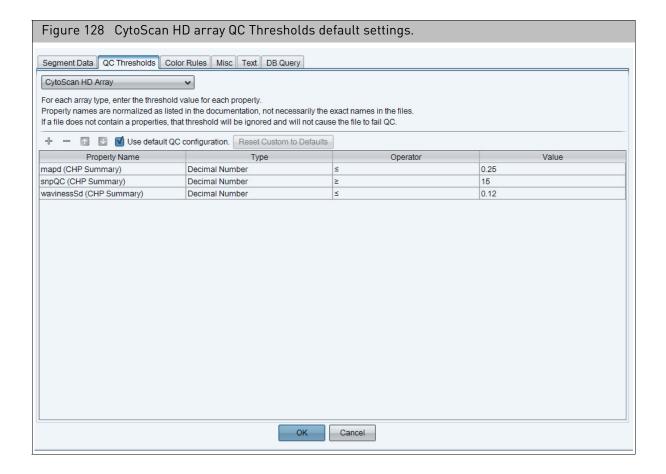
#### **IMPORTANT!** Selecting different QC thresholds is not recommended.

**Note:** When using custom QC thresholds for CytoScan HTCMA in RHAS, these custom thresholds will also need to be updated in ChAS to reflect the desired QC thresholds.

# Viewing QC thresholds

- Click Preferences → Edit User Configuration.
   The User Configuration window appears.
- 2. Click the QC Thresholds tab. (Figure 128)
- 3. Select an array type from the drop-down list.

  Note: QC parameters are specified per array type.



### QC thresholds tab options and descriptions

Option	Description
Property Name	SNPQC is a QC metric for SNP probes that is derived from polymorphic (SNP) probes
	MAPD is a QC metric for all probes used to determine copy number that is derived from both polymorphic (SNP) and non-polymorphic (CN) probes
	<b>Waviness SD</b> is a global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation.
	nd SNP QC is a QC metric for SNP probes that is derived from polymorphic SNP probes in normal diploid regions.
	nd Waviness SD is the same measure as Waviness SD, but only calculates in those regions that are identified as normal diploid.
	Cel Pair Check is a test that inspects each pair of intensity (*.cel) files to determine whether the files have been properly paired and assigned to the correct channel. (OSCHP only)
	<b>DQC (DishQC)</b> measures the amount of overlap between two homozygous peaks created by non-polymorphic probes. DQC of 1 is no overlap, which is good. <b>DQC of 0</b> is complete overlap, which is bad.
	QC Call Rate is the percentage of autosomal SNPs with a call other than NoCall (measured at the Sample QC step).
	SMN MAPD is a QC metric for all probes used to determine copy number that is derived from both polymorphic (SNP) and non-polymorphic (CN) probes calculated during SMN analysis.
	SMN WavinessSD is a global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation calculated during SMN analysis.
	<b>Note:</b> The property names are from the header information of the xxCHP file. QC thresholds are not established for Reproseq Aneuploidy files, but QC metrics are displayed in the QC and Sample Info tab.
Туре	Value or algorithm used for that type of QC.
Operator	The type of comparison performed.
Value	Value assigned to the threshold.

Table 9 Default Copy Number QC Thresholds

Array Type	QC Parameter					
	MAPD	snpQC	Waviness SD	ndsnpQC	ndWavinessSD	
CytoScan 750K and HD Arrays	<u>&lt;</u> 0.25	<u>&gt;</u> 15.0	<u>&lt;</u> 0.12	-	-	
CytoScan XON Arrays	<u>&lt;</u> 0.20	<u>&gt;</u> 10.0	<u>&lt;</u> 0.08	-	-	
CytoScan (Normal Diploid Analysis)	<u>&lt;</u> 0.25	<u>&gt;</u> 15.0	<u>&lt;</u> 0.12	<u>&gt;</u> 15.0	<u>&lt;</u> 0.12	
CytoScan HTCMA	=< 0.28	>= 10	= 0.07	= -	= -	
CytoScan Optima Array MAPD	<u>&lt;</u> 0.29	<u>&gt;</u> 8.5	<u>&lt;</u> 0.12	-	-	
Genome-Wide Human Array SNP 6.0	<u>&lt;</u> 0.35	-	-	-	-	

Table 10 Default Genotyping QC Thresholds

Array Type	QC Parameter			
	DQC	QC Call Rate		
CytoScan HTCMA Arrays	> = 0.88	98.5%		

Table 11 Default SMN QC Thresholds

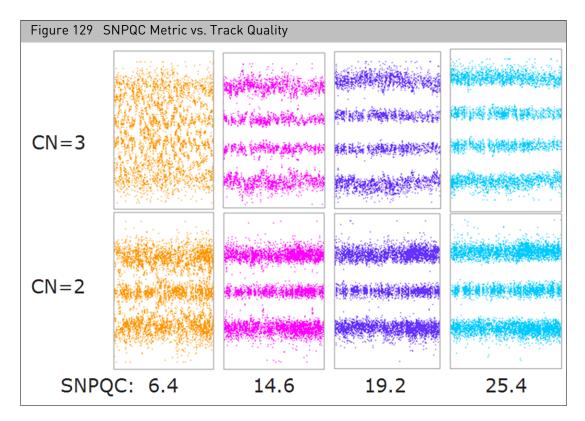
Array Type	QC Parameter			
	SMN MAPD	SMN WavinessSD		
CytoScan HTCMA Arrays	= < 0.35	= < 0.1		

Table 12 Default QC metrics for OncoScan Arrays

Array Type	QC Parameter			
	MAPD	ndSNPQC	CelPairCheckStatus	
OncoScan Arrays	<u>&lt;</u> 0.3	<u>&gt;</u> 26	Pass	

**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. For these sample types, it is recommended using **nd Waviness SD**.

Effect of SNPQC value (for a Single Sample Analysis) on the Allele Difference Track. (Figure 129)



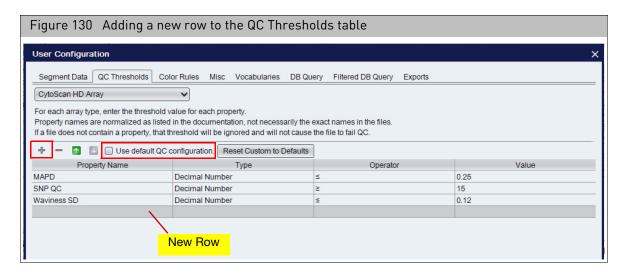
SNPQC is one of the CytoScan within-array QC metrics which provides insight into the overall level of data quality from a SNP perspective. When evaluating the SNPQC values, the key consideration is to ensure that the threshold is exceeded. The quality of the SNP allele data is compromised, from an interpretation perspective, when the SNPQC values are below the recommended acceptance threshold as illustrated by the two left most graphs representing the two and three copy allelic state.

For the CytoScan HD array, when the SNPQC value is below 15, (as illustrated by the data in the two graphs above), the noise within the array is higher than expected. This in turn, compromises the overall data quality and clarity of the 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 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.

# Adding a QC property

- Click Preferences → Edit User Configuration.
   The User Configuration window appears.
- 2. Click the **QC Thresholds** tab. (Figure 130)



- 3. Select an array type from the drop-down list.
- 4. In the QC Thresholds tab, uncheck the **Use default QC configuration** check box.
- 5. Click the **Add** button .A new row appears in the table.

To delete a property row, select the row, then click the **Remove** button.

- 6. Click the **Property Name** field, then enter a new QC property name.
- 7. Click the **Type** field to select **Decimal Number** or **Whole Number**.
- 8. Click the Operator field to select an operator from the drop-down list.
- Double-click the Value field to enter the threshold value for the newly added QC property.
- 10. To add another QC property, repeat steps 5-9.
- 11. Click **OK** to apply the newly added QC threshold(s).

## Editing an existing QC threshold

- 1. Click on an existing **Property Name**, then edit its QC property name.
- 2. Click the **Type** field to edit its **Decimal Number** or **Whole Number**.
- 3. Click the **Operator** field to choose a different operator from the drop-down list.
- 4. Double-click the Value field to edit the current threshold value.
- 5. To edit another existing QC property, repeat steps 1-4.
- 6. Click **OK** to apply the newly edited QC threshold(s).

### Histogram data

## Loading histogram data

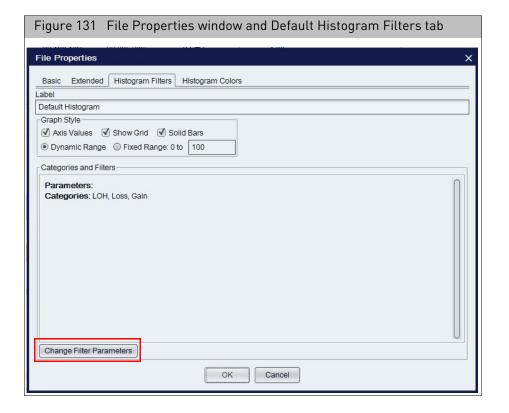
#### IMPORTANT! You must be logged into the ChASDB to view Histogram data.

The histograms load by default. If they are not currently displayed, click  $ChAS\ DB \rightarrow Refresh\ ChAS\ DB\ data$  to view them in the ChAS Browser.

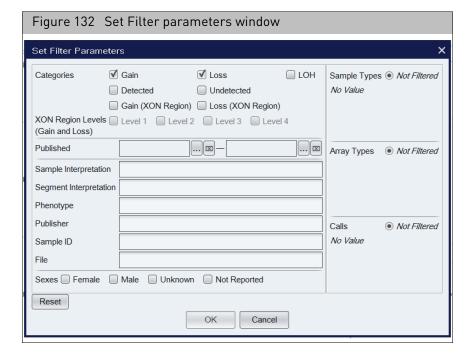
**Note:** The histograms are only available for NetAffx Genomic annotation files for genome build Hg19. The Browser produces an error message, if you try to load Hg19-based histograms while a hg18 or hg38 based NetAffxGenomeAnnotation is currently displayed.

# Changing the default histogram filters

- 1. From the File tree (upper left column), locate the Default Histogram entry.
- Right click on **Default Histogram**.A menu appears.
- Click View/Edit Properties.
   The File Properties window appears. (Figure 131)



- 4. Click the **Histogram Filters** window tab, then use its check box(es), selections, and radio buttons to modify the default Histogram's factory settings.
- 5. Click on the **Change Filter Parameters** button to select new filter settings.
- 6. Optional: To change filter parameters, click **Change Filter Parameters**.



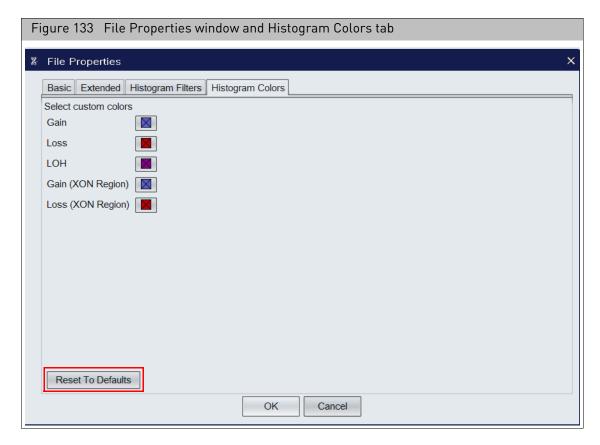
The Set Filter Parameters window appears. (Figure 132)

- 7. Use this window's check boxes, pre-populated entries, and/or the provided text fields (to enter the filter parameters you want).
- 8. Click **OK** to save your changes.

**Note:** If using the aDGV containing segments from both HD and XON arrays, use the Categories filters to limit the histogram data to only HD (check Gain and Loss) or only XON (Gain XON Region and Loss XON region)

# Changing the default histogram colors

- 1. From the File tree (upper left column), locate the Default Histogram entry.
- Right click on **Default Histogram**.A menu appears.
- Click View/Edit Properties.The File Properties window appears.
- 4. Click on the **Histogram Colors** tab. (Figure 133)

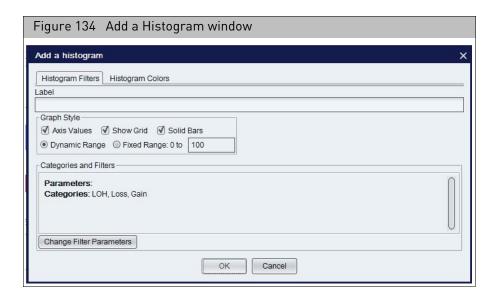


- 5. Click on the color square representing the histogram track you want to change. A change color window appears.
- 6. Select a new color, then click **OK**.
- 7. Repeat steps 5-6 as needed.
- 8. Click **OK** to save your color changes or click **Reset to Defaults** to return to the default color settings.

# Adding filtered histogram data

You can add additional histograms filtered on certain properties for viewing in the Detail View.

Go to ChASDB → Add Histogram
 The Add a Histogram window appears. (Figure 134)



- 2. In the Label text box, enter a name for the histogram.
- 3. In the **Graph Style** pane, click the appropriate check box(es) and/or radio button(s) to define your filtered Histogram's graphic style.
- 4. Use this window's check boxes and radio buttons to create the desired graph style. Click on the **Change Filter Parameters** button to select filters to be applied to your new histogram. Click on the **Histogram Colors** tab to select colors for your new histogram tracks. See "Changing the default histogram colors" on page 141 for details.
- 5. Click OK.

The histogram is added to the bottom of the File tree list. The histogram can be moved to a different position in the Detail View by clicking on the name of the histogram, then dragging it to the desired location within the File tree.

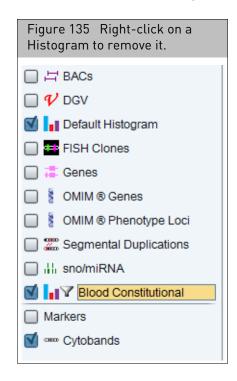
You can access what segments are in a particular bin of the histograms by right-clicking on the histograms and selecting, **Show Histograms items in bin**. For more details, see "Segment intersections" on page 391.

**Note:** When right clicking to Show Histograms in bin, the percentages for DB Coverage Count and DB Overlap Count columns in the Segment Intersections window represent the percentage intersection with the query region (the light blue vertical bars) shown in the Segment Intersections window. For more details on this Segment Intersection window, see Figure 426 on page 392.

# Removing a histogram

- 1. Locate the Histogram you want to remove in the Files window pane. (Figure 135)
- 2. Right-click on it, then click **Close**.

**Note:** Closing a Histogram removes the Histogram from the Detail View. In order to view this Histogram again, you must recreate it using the Add Histogram steps listed above. See "Adding filtered histogram data" on page 143.



# 7

### Viewing Data

#### This chapter includes:

- "Displaying options of analysis results data"
- "Overview of ChAS window components" on page 146
- "Files list" on page 147
- "Data types list" on page 149
- "Named settings" on page 150
- "Status bar" on page 150
- "Display area" on page 151
- "Changing the NetAffx genomic annotation file version" on page 153

### Displaying options of analysis results data

- Graphic Displays
  - See "Displaying Data in Graphic Views" on page 154.
- Tables
  - See "Displaying data in table views" on page 327.

#### 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" on page 219.

- 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" on page 268.

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

See "Using the overlap map and filter" on page 280.

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

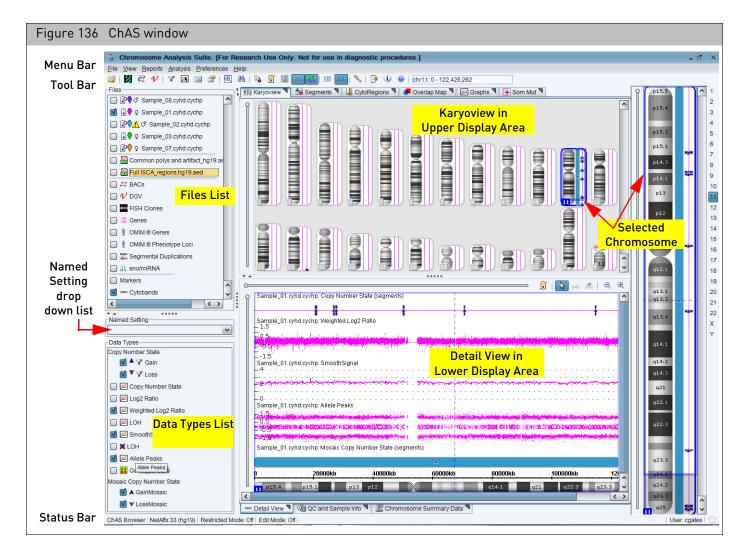
See "Creating and editing AED files" on page 288.

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

See "Exporting results" on page 413.

• Save setups of ChAS for different tasks in user profiles and named settings. See "User profiles and named settings" on page 436.

### Overview of ChAS window components



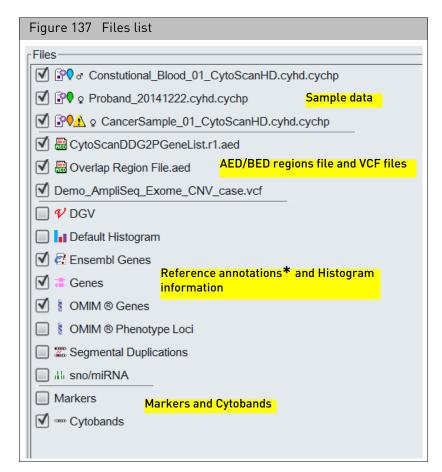
The ChAS window components (Figure 136) include:

- **Menu Bar**: Access to the functions of the software.
- Tool Bar: Quick access to commonly used functions.
- Files List: Displays data and annotation files that can be displayed. (page 147)
- Data Types List: Displays the type of data available in the files. (page 149)
- Named Settings: Displays a list of the previously saved display settings for ChAS. (page 150)
- Status Bar (page 150): Displays:
  - Software status
  - ChAS Browser's NetAffx Genomic Annotation file version.
  - hg version
  - The annotation or probe that your mouse pointer is nearest (in the Detail View).

- Display Area (page 151): Displays the following data in graphical and table formats:
  - Analysis results graph data.
  - Detected Segments
  - Histograms
  - Regions
  - Reference Annotations

### Files list

The Files list (Figure 137) 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.



<sup>\*</sup>Reference annotations displayed may be different than shown above depending on the NetAffx Genomic Annotation file loaded.

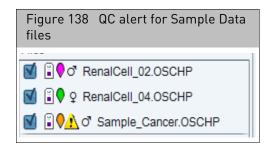
Displayed files grouped by type:

• Sample Data

Colored nibs  $\sqrt{\phantom{a}}$  display the color used for the data lanes for that sample in the Karyoview, Selected Chromosome View, and Detail View. The appropriate gender symbol is displayed to the right of the colored nib.

If a loaded file has a QC parameter that is out of range, an alert symbol 🔼 appears next to the file name. (Figure 138)





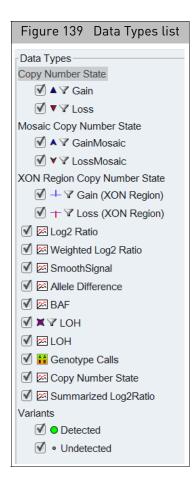
- Region Information Files Icons indicate the file type (AED or BED) and whether the loaded files have been selected as a CytoRegion file 🛄 or Overlap Map 💋 file.
- Reference Annotations (loaded during software start-up). **Note:** Icons indicate the annotation type.
- Histogram information (loaded during software start up if connected to ChAS DB).
- Cytobands (Separated from other reference annotations because cytobands cannot be moved in the displays)

See "Selecting data for display" on page 192. 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" on page 318..

### Data types list

The Data Types list (Figure 139) 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" on page 119...



The Data Types list enables 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  $\mathbf{Y}$  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" on page 172 for more information.

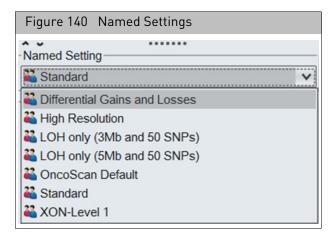
Unselected data is also concealed from the different tables and graphs.

See "Selecting data types for display" on page 194. for information on using the Data Types list to select different data types for display.

### Named settings

The Named Settings drop-down list (Figure 140) 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. See "Using restricted mode" on page 276.



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

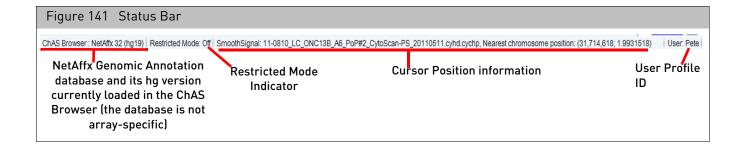
See Chapter 20, "User profiles and named settings" on page 436 for information on creating and using Named Settings.

**Note:** The OncoScan Default Named Setting has no filters applied so all segments called by the TuScan algorithm can be viewed. Users can create their own appropriate custom filter Named Setting for their data, see "User profiles and named settings" on page 436.

### Status bar

The status bar (Figure 141) (very bottom of browser) displays information on:

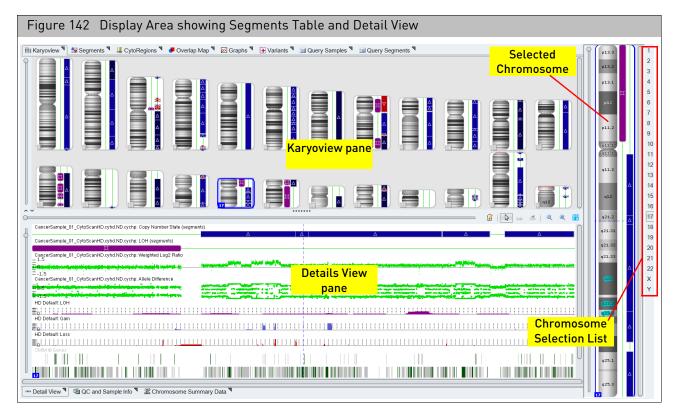
- NetAffx Genomic Annotation database and its hg version
- Restricted Mode status (See "Using restricted mode" on page 276.)
- Edit Mode status (See "Using edit mode" on page 226.)
- Cursor (Mouse Over) Position
- User Profile ID



### Display area

The Display Area (Figure 142) is divided into three panes:

- "Upper panes" on page 152
- "Selected chromosome view" on page 152
- "Lower panes" on page 152



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.

#### Upper panes

The Upper pane displays the following data in graphics and table formats:

- "Karyoview" on page 155: Displays selected segment types for selected sample files for all chromosomes.
- "Segments table" on page 337: Displays a list of the detected segments in the selected sample files.
- "CytoRegions table" on page 273: 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.
- "Viewing the overlap map table" on page 284: 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" on page 345: Displays marker data for the loaded and selected xxCHP files.
- "Variants table" on page 351: Displays somatic mutation data from OncoScan CNV Plus and Carrier Variant data from CytoScan HTCMA arrays.
- Query Samples Table: Displays sample level results from searches performed on the ChAS DB.
- Query Segments Table: Displays segment level results from searches performed on the ChAS DB.

### Selected chromosome view

The Chromosome View displays detected segments in selected sample files for the chromosome selected in the Karyoview, while the Chromosome Selection List (far right column) displays its number.

See "Selected chromosome view" on page 170 for more information.

#### Lower panes

The Lower Pane displays:

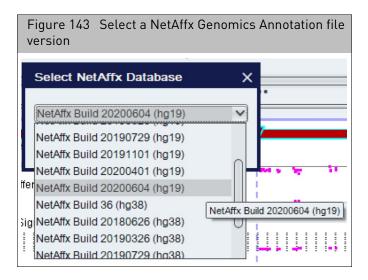
- "Detail view" on page 172: Displays the selected section of the chromosome displayed in Selected Chromosome View and includes:
  - Detected segments, variants, and graph data in selected xxCHP files.
  - Histograms, AED/BED file regions and annotations.
  - Reference annotation files
- "QC and sample info tab" on page 355: Displays QC metric information as well as information about the loaded Sample and Region (AED) files.
- "Chromosome summary data" on page 362:
  - 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).
  - Calculated Properties: Calculates the Percent Autosome LOH in the sample based on the LOH filter setting.

### Changing the NetAffx genomic annotation file version

- At the top menu bar, click File → ChAS browser NetAffx Genomic Annotation file version.
- 2. From the **Select NetAffx Database** drop-down menu (Figure 143), click to select the NetAffx Build version you want, then click **OK**.

**Note:** If there are loaded xxCHP files with hg version different from the selected NetAffx database, a message appears stating that these data files will be closed (in both the ChAS browser and the MSV) <u>before</u> the NetAffx annotations are loaded.

**Note:** More current NetAffx Genome Annotation files have the date the file was generated imbedded into the file name. To load the most current file from your library folder, select the file with the most recent date.



For more information on what types of data is in the database files and how this information varies in different versions, see "ChAS browser NetAffx Genomic Annotations" on page 486.



### Displaying Data in Graphic Views

#### This chapter includes:

- "Graphic views" on page 154
- "Controlling the display of data" on page 191
- "Edit configurations in the misc tab" on page 187
- "Learning more about features" on page 205

ChAS provides multiple graphic views for detected segments and other data.

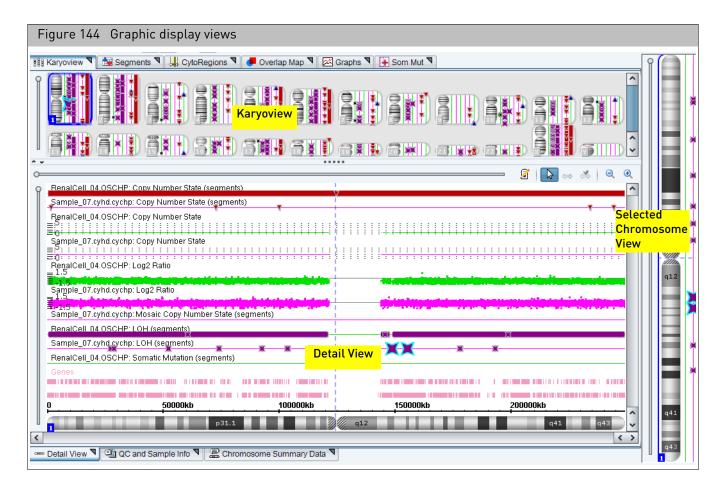
Use these graphic views to:

- Get an overview of the detected segments.
- Get an overview of segments stored in the ChAS DB.
- Compare segments between samples.
- View sample data for the whole genome.
- 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.

### **Graphic views**

Data can be displayed in the following graphic views: (Figure 144)

- "Karyoview" on page 155
- "Whole genome view" on page 159
- "Selected chromosome view" on page 170
- "Detail view" on page 172



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

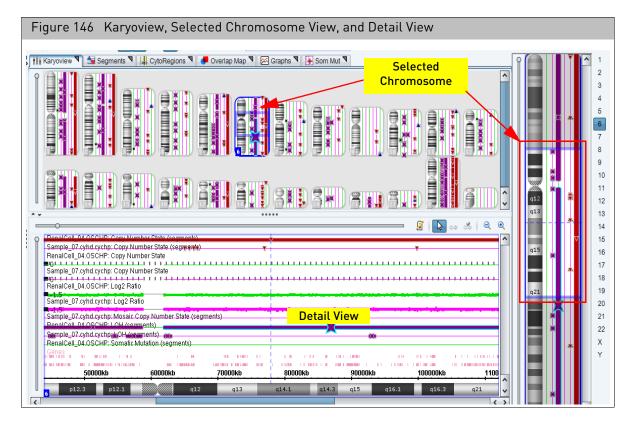
The Karyoview (Figure 145) displays a genome-wide view of the detected segments and other data.

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.



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 146)



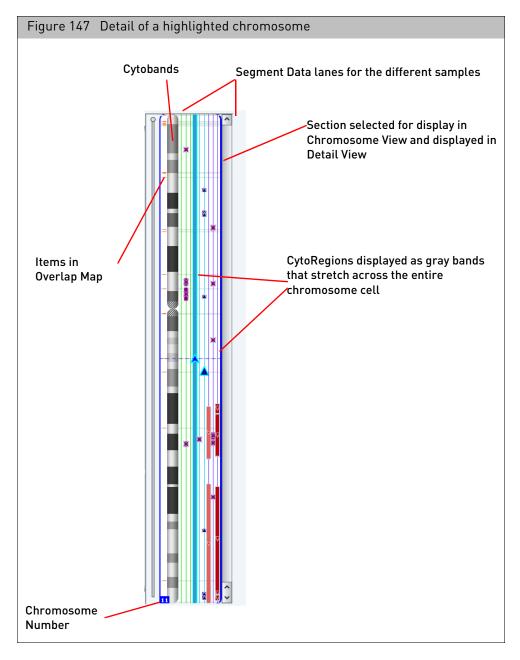
**Note:** To easily remove the blue highlight surrounding the selected chromosome for image captures, go to  $View \rightarrow Hide Karyoview Highlights$ .

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 detailed below:

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

The following information is displayed for each chromosome: (Figure 147)

- Chromosome number
- Cytobands
- Segment Data Uses 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. See "Changing the grouping of samples and data types" on page 195.
  - Change the position of samples. See "When the lanes are grouped by sample, the different segment types for each sample are kept together in the Karyoview and Selected Chromosome View (Figure 188) and in the Detail View." on page 196 and "Lanes grouped by sample" on page 196.
- Variant Data (only available for OSCHP and RHCHP)
- CytoRegions in selected CytoRegion File.
- Items in selected Overlap Map file.



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.

If you assigned an Overlap Map file, the overlap map items are displayed as small rectangles to the left of the cytobands. Its color is the same color used in the Details View.

**Note:** 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. Also, you can right-click on a feature in the Karyoview or Selected Chromosome View to open a shortcut menu of options, as shown in Figure 148.



### Whole genome view

The Whole Genome View (Figure 149) displays the following data in a single display from Chromosome 1 to Y for:

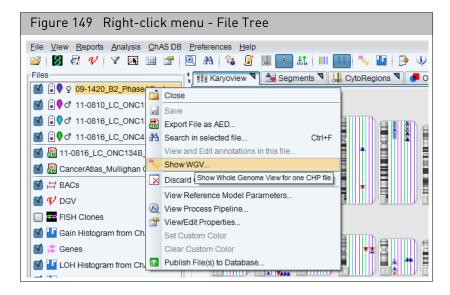
- CYCHP: Log2 Ratio, Weighted Log2 Ratio, Copy Number State, Smooth Signal, LOH Allele Difference, and B-allele Frequency.
- XNCHP: Log2 Ratio, Weighted Log2 Ratio, Smooth Signal, LOH, Allele Difference, and B-allele Frequency.
- CNCHP: Log2 Ratio, Copy Number State, Smooth Signal, Allele Difference
- RHCHP: Log2Ratio, Smooth Signal, Copy Number State, LOH, Allele Difference, and B-allele Frequency.
- OSCHP: Log2 Ratio, Weighted Log2 Ratio, Copy Number State, Smoothed Signal, Allele Difference, and B-allele Frequency.
- ReproSeq Aneuploidy .zip: Copy number state data provides the sequencing tile data

**Note:** When opening ReproSeq for the first in the Whole Genome View, the graphs may appear empty. Use **Choose Data** to select the Copy Number Graph type to view the data.

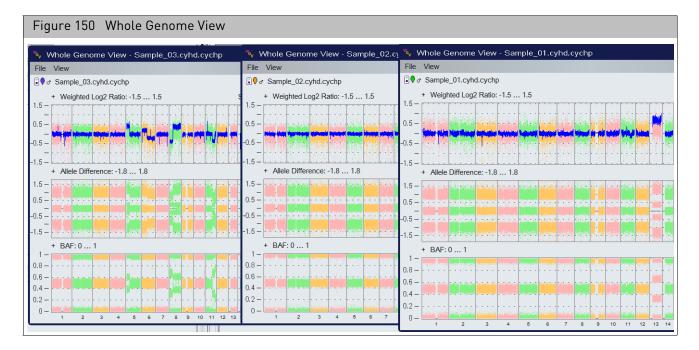
**Note:** The Y axis on the left represents the data points in that graph (Log2/Weighted Log2 Ratio). The Y axis on the right represents the line graph data (Smooth Signal/Copy Number).

# Displaying the Whole Genome View for a sample(s)

 Right-click on the Sample name in the File Tree (Figure 149), then select Show WGV.



2. The Whole Genome View for the selected file(s) opens. (Figure 150)

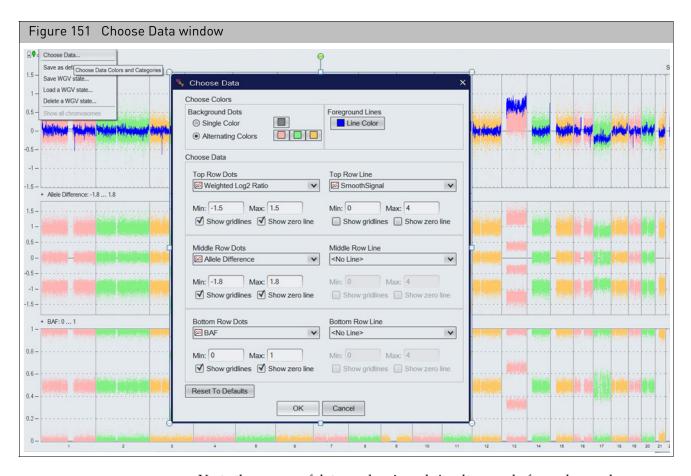


- 3. To select a different sample(s) to display:
  - Select File → Choose File.
  - Click on a different file name, then click **OK**.

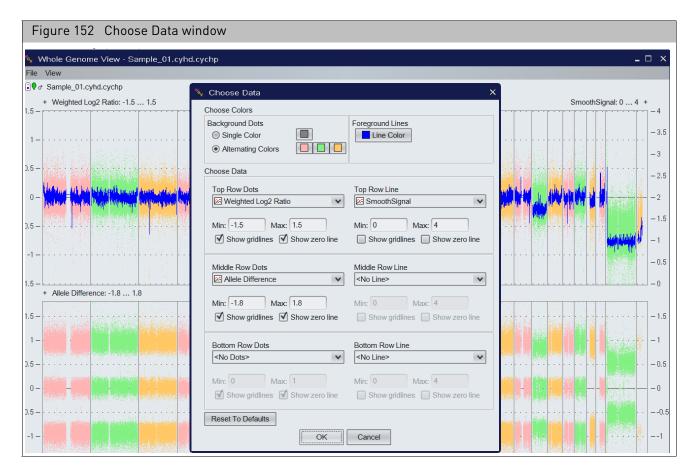
### Changing graph types

1. Click on View  $\rightarrow$  Choose Data.

The Choose Data window appears. (Figure 151)



- Up to three rows of data can be viewed simultaneously for each sample.
- Use the Row Data and/or Row Line to assign which data tracks to view.
- Select <no Dots> or < no Line> to disable the display (of any data) in the row.
   Figure 152 is an example of no data being assigned to the Bottom Row, therefore only two rows of data are displayed in the Whole Genome View.

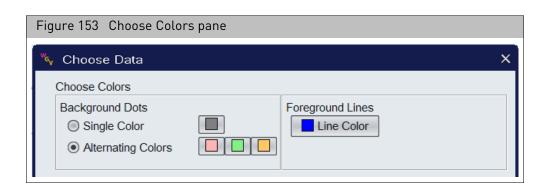


**Note:** The Y axis settings for the Whole Genome View are initially determined from the Y axis settings set in the Detail View Graph Settings. The Y axis on the left of the WGV pertains to the Row Data. The Y axis on the right of the WGV pertains to the Line data. Also, not all graphs are available for a given xxCHP file. If a graph type is selected in which data is not available, the graph will appear with no data points.

#### Changing colors of the data points or line data

Graphs can be viewed in a single color or alternating colors every three chromosomes.

1. Click View → Choose Data. (Figure 153)

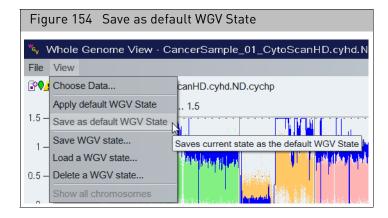


- 2. Select the radio button to change to either 1 color data points or alternating 3 color data points.
- 3. Click on the colored square(s) to open a Color Selection palette. Select new colors for the Whole Genome View.
- 4. Click **OK** to return to the Whole Genome View with your new selection.

**Note:** Changing Data display and/or Colors affects the current sample only. To save these settings as default setting (whenever a file is opened), save these settings as the Default WGV State. (See "Setting a default WGV state display and colors".)

## Setting a default WGV state display and colors

- 1. Select the Data and Colors to be displayed as your Default settings when opening files in the WGV, as described in "Changing colors of the data points or line data" on page 162.
- 2. Select View  $\rightarrow$  Save as default WGV State. (Figure 154)



A confirmation message appears.

3. Acknowledge the message, then click **OK** to save these settings as the Default WGV display.

**Note:** If changes have been made to colors, data or Y axis values, you can return to this WGV Default State by clicking View → Apply default WGV State or click Load a WGV State → Default.

### Creating WGV states

Multiple WGV States can be created, then saved for a quick selection of different graph/color settings.

- Select the Data and Colors to be displayed as your Default settings when opening files in the WGV.
- 2. To save these selections as a WGV State, click on  $View \rightarrow Save WGV State$ .
- 3. Enter a name in the dialog box, then click **OK** to save this new WGV State.

### Applying a saved WGV state(s)

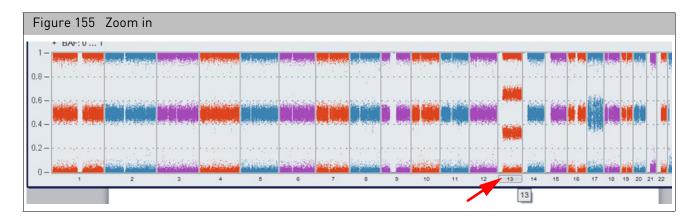
- 1. Select View  $\rightarrow$  Load a WGV State.
- 2. Choose a saved WGV state from the drop-down list.
- 3. Click **OK** to apply the WGV State (or click **Cancel** to keep the current display).

### Deleting a saved WGV state

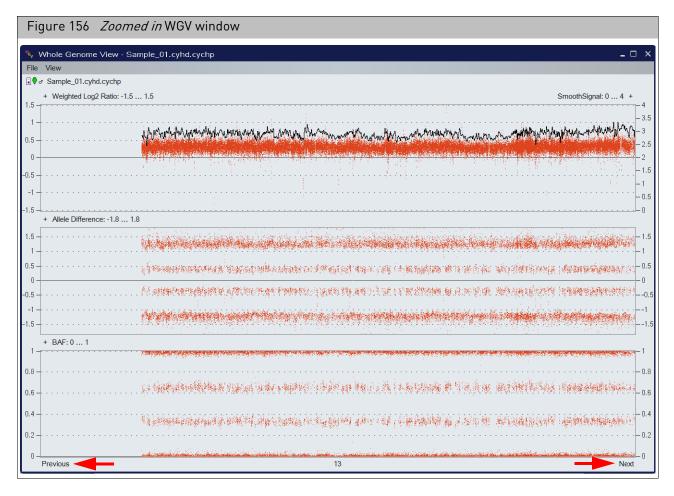
- 1. Click on View → Delete a WGV State.
- 2. Select the saved WGV state you want to delete from the drop-down list, then click **OK**.
  - A message appears asking if you are sure you want to delete it.
- 3. Click **OK** or click **Cancel** to keep the WGV State.

### Using the WGV zoom feature

1. Click on the chromosome number that appears along the bottom row, as shown in Figure 155.



2. The following "zoomed in" WGV window appears. (Figure 156)

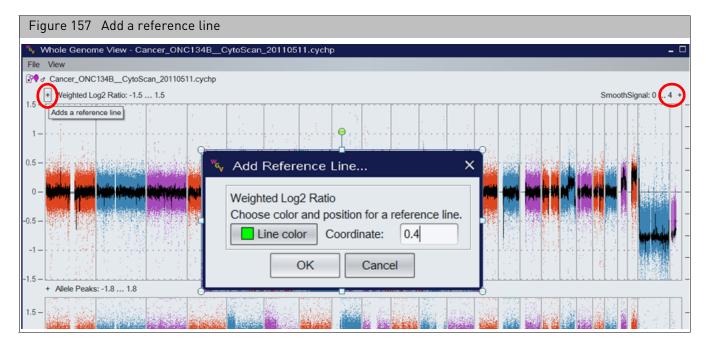


- 3. Click **Previous** or **Next** to view adjacent chromosomes in this window.
- 4. To return to the Whole Genome View (Figure 155), click on the chromosome number again or click View→ Show all chromosomes.

### Adding a reference line to the WGV

1. Click on the + (above the panel) to add a reference line to that panel.

**Note:** When a row of data has both Dots and Line data, click the + sign next to the data type to which you would like to add the Reference Line.



- 2. Click on the **Line color** button to select a color for the Reference Line.
- 3. In the **Coordinate** text field, enter an approximate coordinate (based on the Y axis) where you want the Reference Line placed.

**Note:** The Reference Line can be dragged and dropped to a different location once placed onto the graph.

- 4. Click OK.
- Repeat steps 2-4 to place additional Reference Lines onto the graph.
   Note: Additional Reference Lines can also be added by right-clicking in the graph and selecting Add Reference Line for... (Figure 158)
- 6. Delete a single Reference Line by right-clicking on the Line and selecting **Delete Selected Line**. Remove all Reference lines by right-clicking in the graph and selecting **Delete All Lines**. (Figure 158)
- 7. To change the color or coordinate of a Reference Line, right-click on the selected Reference Line, then select **Edit Selected Line**. (Figure 158)



### Adding a comparison file

Use this feature to view two samples in the same WGV window.

**Note:** Before using this feature, both analysis files MUST BE loaded and available in the ChAS Browser.

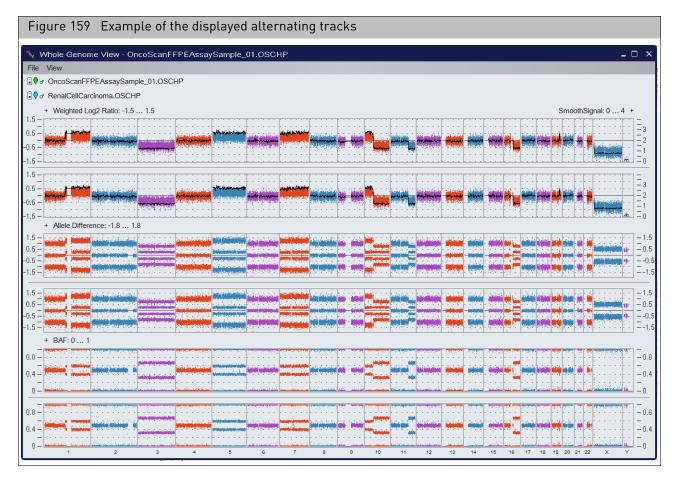
- 1. Open the first file in the Whole Genome View, as you normally would.
- 2. Click File  $\rightarrow$  Add Comparison File.

A Comparison File window opens with available analysis results files loaded into ChAS Browser.

3. Click to highlight the file to be viewed with the first file (already open in the WGV), then click **OK**.

The data for the files are loaded in the same window (Figure 159) displaying the following alternating tracks:

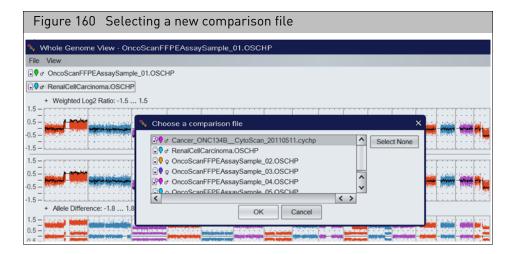
- Track 1: Weighted Log 2 and Smooth Signal for File 1
- Track 2: Weighted Log 2 and Smooth Signal for File 2
- Track 3: Allele Difference for File 1
- Track 4: Allele Difference for File 2
- Track 5: B-allele Frequency for File 1
- Track 6: B-allele Frequency for File 2



**Note:** To remove a comparison file, click **File**  $\rightarrow$  **Remove Comparison File**.

### Selecting a new comparison file

- 1. Click on the second of the filenames to open the file selection window
- 2. Highlight a new file from the **Choose a comparison file** window (Figure 160), then click **OK**.



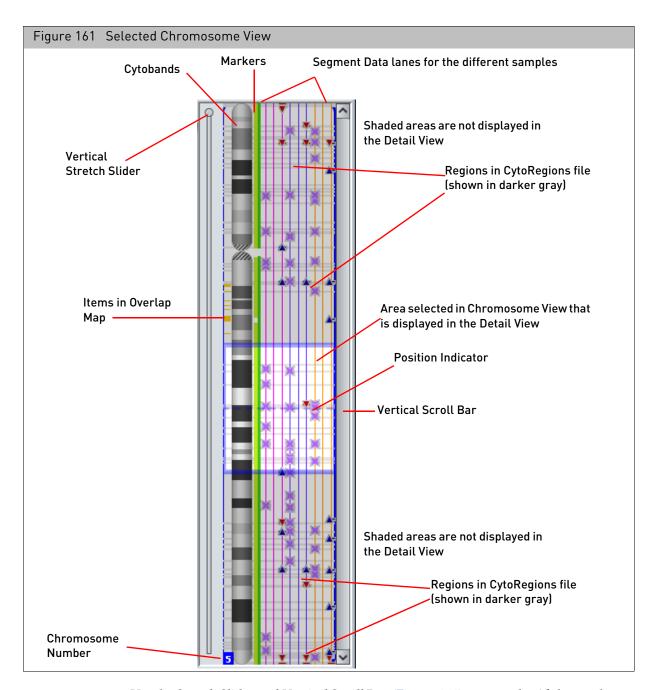
The newly selected file and its data is displayed.

#### Exporting a WGV

1. Click File  $\rightarrow$  Export window PNG.

### Selected chromosome view

The Selected Chromosome View (Figure 161) is similar to the Karyoview, but it displays a single selected chromosome at higher magnification. Click, then drag in the Chromosome View to select an area for display in the Detail View.



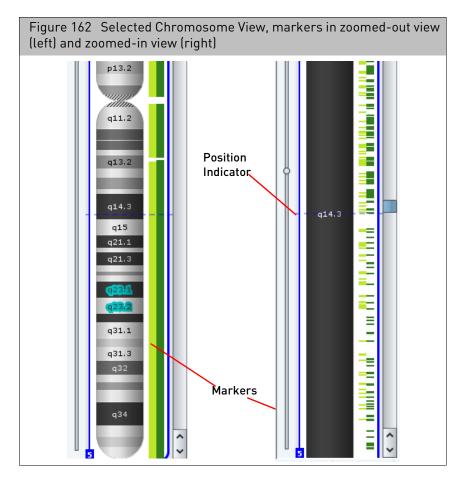
Use the Stretch Slider and Vertical Scroll Bar (Figure 161) or press the **Alt** key and turn the mouse wheel to zoom in on a section of the **Selected Chromosome View**.

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).

If you have selected an Overlap Map file, the overlap map items are displayed as small rectangles (Figure 161) to the left of the Cytobands. Its color is the same color used in the Details View.

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.

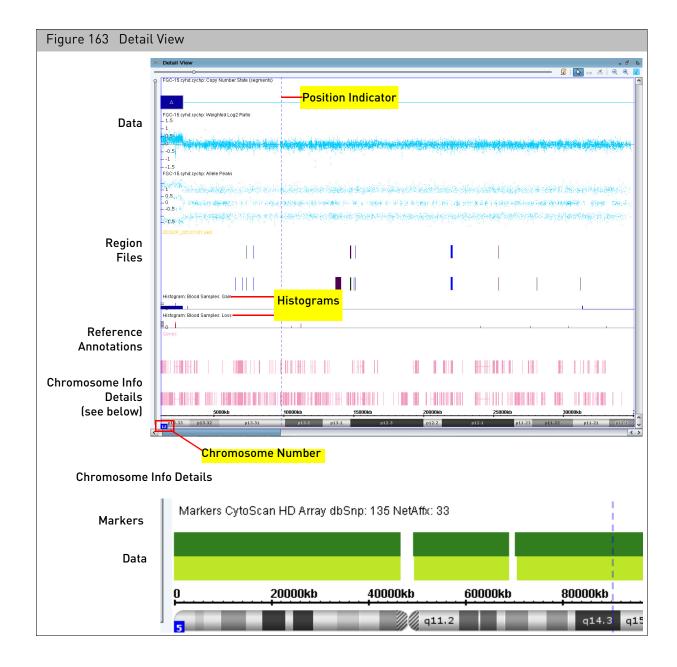
The marker positions are displayed to the right of the Cytobands. When zoomed out, they appear as green ribbons. When zoomed in, the markers and their positions can be seen, as shown in Figure 162.



- SNP markers are displayed in the light green band nearest the cytobands. The SNP marker/probe names in the CytoScan start with the letter 'S'.
- There is one marker track for every distinct array type that is loaded.
- Copy Number markers are displayed in the dark green band nearest the detected segments. The non-polymorphic copy number probe names on the CytoScan 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 "Selected chromosome view" on page 170.

### **Detail view**

The Detail View (Figure 163) enables you to look in detail at the detected segments, marker data, regions, and reference annotations for the loaded files.



#### Viewing CytoScan XON data in detail view

CytoScan XON data is categorized into four different levels.

- Level 1: Includes genes with the highest level of evidence: developmental delay, epilepsy, ASD, XLID, Metabolic disorders, hereditary cancer OMIM<sup>™</sup> Morbid genes.
- Level 2: ClinVar genes not covered in Level 1.
- Level 3: Other OMIM genes.
- Level 4: Other Ref Seq, UCSC, Ensembl genes, LOVD.

In the Filters tab, when only Level 1 is selected and Levels 2-4 remain unchecked (as shown in Figure 164 on page 174) then:

XON Region segments that overlap regions of the genome designated as Level 1 is visible in the XON Region Segment Track and in the Segments Table. All remaining data (log2 ratio, weighted log2 ratio, smooth signal, allele difference, B-allele Frequency) contained within Level 1 regions is colored the same color as the color nib of the sample. XON Region segments that overlap regions of the genome designated as Levels 2-4 are not visible on the XON Region segment track or in the Segments Track. All remaining data (log2 ratio, weighted log2 ratio, smooth signal, allele difference, B-allele Frequency) contained within Level 2-4 regions is colored gray.

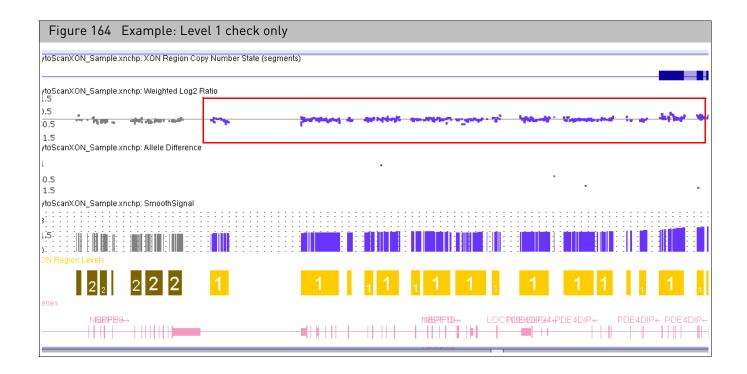
In the Filters tab, when Levels 1 and 2 are selected and Levels 3 and 4 remain unchecked (as shown in Figure 165 on page 174) then:

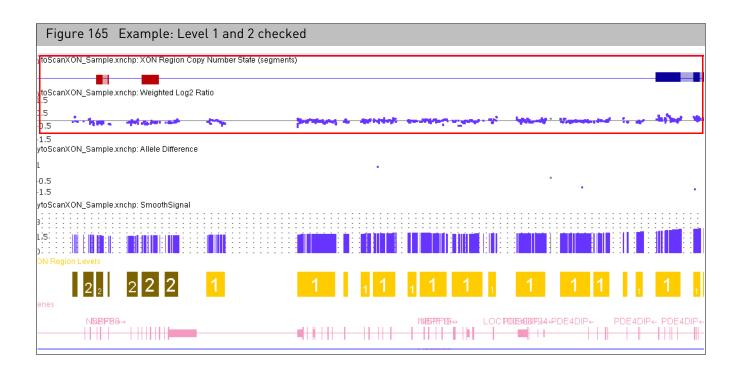
XON Region segments that overlap regions of the genome designated as Level 1 or Level 2 will be visible in the XON Region Segment Track and in the Segments Table. All remaining data (log2 ratio, weighted log2 ratio, smooth signal, allele difference, Ballele Frequency) contained within Levels 1 or 2 regions will be colored the same color as the color nib of the sample. XON Region segments that overlap regions of the genome designated as Levels 3 or 4 will not be visible on the XON Region segment track or in the Segments Table. All remaining data (log2 ratio, weighted log2 ratio, smooth signal, allele difference, B-allele Frequency) contained within Level 3 and 4 regions are colored gray.

If selected Level 3, then XON Region Segment calls are displayed on the XON Region Segment track and the marker level data are colored the same color as the color of the color nib of the sample.

When all four Levels are selected in the Filters Tab, all XON Region segment calls for the sample are displayed and all marker level data is colored the same color as the color of the color nib of the sample.

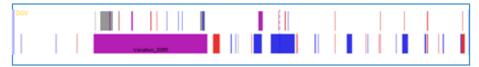






### Annotation color codes

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

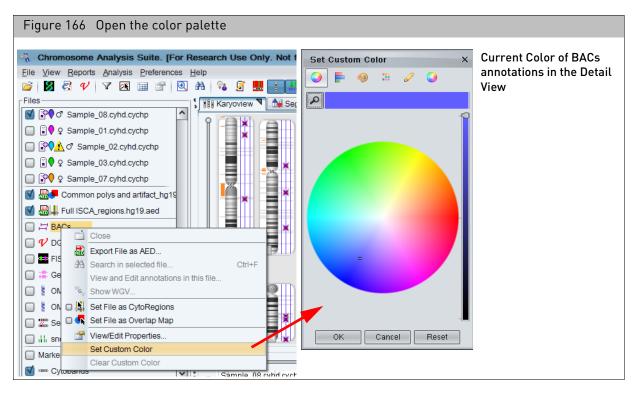
### Annotation OMIM color codes

In Browser annotation files version NA32.3 and higher, the following OMIM colored gene entries were generated by genome.ucsc.edu and are based on the associated OMIM phenotype map key. For more information on OMIM display conventions, go to: www.genome.ucsc.edu

- Lighter Green for phenotype map key 1 OMIM records the disorder has been placed on the map based on its association with a gene, but the underlying defect is not known.
- Light Green for phenotype map key 2 OMIM records the disorder has been placed on the map by linkage; no mutation has been found.
- Dark Green for phenotype map key 3 OMIM records the molecular basis for the disorder is known; a mutation has been found in the gene.
- Purple for phenotype map key 4 OMIM records a contiguous gene deletion or duplication syndrome; multiple genes are deleted or duplicated causing the phenotype.
- Light Gray for Others no associated OMIM phenotype map key info available.

#### Changing an annotation color

1. Right-click an annotation type in the Files window pane and select **Set Custom Color** on the shortcut menu, as shown in Figure 166.



- 2. Specify a color for the selected annotation type using the color controls in the color palette (Figure 166), then 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 types of data for CytoScan (CYCHP):

Table 13 Data for CytoScan array

Data Types	Definition	
Detected Segment Types	Detected Segment Types	
GainMosaic	Non-integer amplifications or duplications	
LossMosaic	Non-integer hemizygous or homozygous deletions	
Gain	Amplifications or duplications	
Loss	Hemizygous or homozygous deletions	
LOH	Loss of Heterozygosity	
Probe array data (displayed as graph data)		
Copy Number State	HMM-derived integer Copy Number State	
Log2 Ratio	Per marker Log2 Ratio of normalized intensity with respect to a reference, with further correction for sample specific variation.	
Weighted Log2 Ratio	Contains the Log2 Ratios processed through a Bayes wavelet shrinkage estimator. These processed values are input to the CNState algorithm HMM.	
LOH	Loss of Heterozygosity	
Allele Difference	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.	
Genotype Calls	SNP genotype calls (single sample, BRLMM-P-plus algorithm)	
Smooth Signal	Gaussian Smoothed Calibrated Copy Number Estimate	
B-allele Frequency	Number of B alleles/number of A+B alleles used to show allelic imbalances.	

The Detail View displays the following types of data for a CytoScan HTCMA array:

Table 14 Data for CytoScan HTCMA array

Data Types	Definition
Detected Segment Types	
Gain	Amplifications or duplications
Loss	Hemizygous or homozygous deletions
LOH	Loss of Heterozygosity
Probe array data (displayed as graph data)	
Copy Number State	HMM-derived integer Copy Number State

Table 14 Data for CytoScan HTCMA array

Data Types	Definition
Log2 Ratio	Per marker Log2 Ratio of normalized intensity with respect to a reference, with further correction for sample specific variation.
LOH	Loss of Heterozygosity
Allele Difference	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.
Genotype Calls	SNP genotype calls (single sample, BRLMM-P-plus algorithm)
Smooth Signal	Gaussian Smoothed Calibrated Copy Number Estimate
B-allele Frequency	Number of B alleles/number of A+B alleles used to show allelic imbalances.
Variants	Location and detection of variants.

**Note:** 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 (on the CytoScan HD 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 into account. All non-polymorphic (copy number) and the vast majority of SNP probes are NOT affected by this change, and will continue to display all graphs and their data points from the CytoScan HD Array CYCHP files.

Table 15 Data for CytoScan XON Array

Data Types	Definition	
Detected Segment Ty	Detected Segment Types	
XON Region Gain	Amplifications in XON regions	
XON Region Loss	Hemizygous or homozygous deletions in XON regions	
LOH	Loss of Heterozygosity	
Probe array data (displayed as graph data)		
Log2 Ratio	Per marker Log2 Ratio of normalized intensity with respect to a reference, with further correction for sample specific variation.	
Weighted Log2 Ratio	Contains the Log2 Ratios processed through a Bayes wavelet shrinkage estimator. These processed values are input to the CNState algorithm HMM.	
LOH	Loss of Heterozygosity	

Table 15 Data for CytoScan XON Array

Data Types	Definition
Allele Difference	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.
Genotype Calls	SNP genotype calls (single sample, BRLMM-P-plus algorithm)
Smooth Signal	Gaussian Smoothed Calibrated Copy Number Estimate
B-allele Frequency	Number of B alleles/number of A+B alleles used to show allelic imbalances.

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

Table 16 Data for Genome-Wide SNP Array 6.0 (CNCHP)

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

Table 17 Data for OncoScan Array (OSCHP)

Data Types	Definition
Detected Segment Types	
Gain	Amplifications or duplication
Loss	Hemizygous or homozygous deletions
LOH	Loss of Heterozygosity
Probe array data (displayed as graph data)	
Copy Number State	TuScan derived Copy Number State.
Log2 Ratio	Per marker Log2 Ratio of normalized intensity with respect to a reference, with further correction for sample specific variation.
Weighted Log2 Ratio	Contains the Log2 Ratios processed through a Bayes wavelet shrinkage estimator. These processed values are input to the CNState algorithm HMM.
LOH	Loss of Heterozygosity.
Allele Difference	Difference of A signal and B signal, each standardized with respect to their median values in the reference.
Smooth Signal	Smoothed Calibrated Copy Number Estimate.
Variants	Location and detection of Somatic Mutation. (OncoScan FFPE Assay only)
B-allele Frequency	Number of B alleles/ number of A+B alleles, used to show allelic imbalances.

Table 18 ReproSeq Aneuploidy data

Data Types	Definition
Detected Segment Types	
Gain	Amplifications
Loss	Hemizygous or Homozygous deletions
Graph Data	
Copy Number State	The copy number state of sequence tiles

See "Changing graph appearance" on page 198 for more information about controlling the display of graph data.

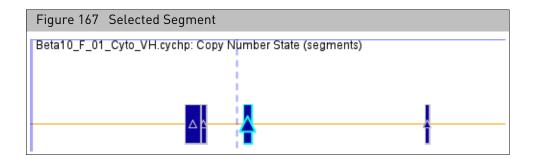
In addition, the Detail View displays:

- 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, structural variants, and so forth.

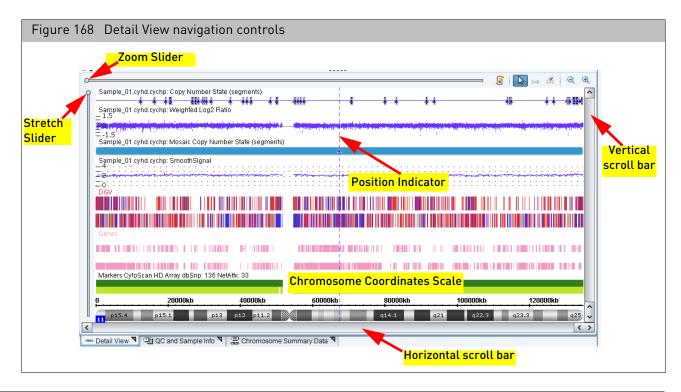
You can expand or contract the annotations. See "Expanding and contracting annotations" on page 197.

- Database Display
  - Default Histograms: displays all segments in the database.
  - Filtered Histograms: displays all segments meeting the filter criteria set by the user. For information on how to create a filtered histogram, see "Adding filtered histogram data" on page 143.
- 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 167)



### Navigation controls in detail view



Control	Function
Chromosome Coordinates scale	Shows the position along the genome.
Zoom Slider	Controls the horizontal zoom and the area of the chromosome displayed.
3x Zoom In	Press Ctrl + Plus to view up to three preset Zoom In settings.
3x Zoom Out	Press Ctrl + Minus to view up to three preset Zoom Out settings.
Zoom in Using Click and Drag	Place your mouse cursor over a point of interest. Press Shift while holding down the left mouse button. Drag the mouse cursor to frame/zoom in on your point of interest.
Stretch Slider	Controls the vertical stretch of the Display area.
Scroll bars	Used to select the area displayed after zooming or stretching the vertical or horizontal scale.
Position Indicator	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.

Use 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

# Obtaining summary metrics for a zoomed in region

You can obtain summary metrics from the data tracks displayed in the Detail View once you have zoomed in to a region of interest.

- 1. In the Detail View, zoom in to a region of interest using any of the techniques described above.
- 2. Hover the mouse over a data point on the data track for which you would like summary metrics. For example, log 2 ratio, weighted log 2 ratio, Allele Difference, etc.

A pop-up appears and displays the following metrics for markers currently displayed in the Detail View:

- Region Size
- Number of Markers
- Min Marker Value
- Max Marker Value
- Mean Value
- Standard Deviation
- Median Value

#### Selecting a chromosome section for display

Data from the same sample files is displayed in all 3 views, at different scales.

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

- "Karyoview and selected chromosome view"
- "Coordinate range box" on page 184
- "Zooming to a selected item" on page 185
- "Navigation controls in detail view" on page 182

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

#### Selecting a chromosome for detailed examination

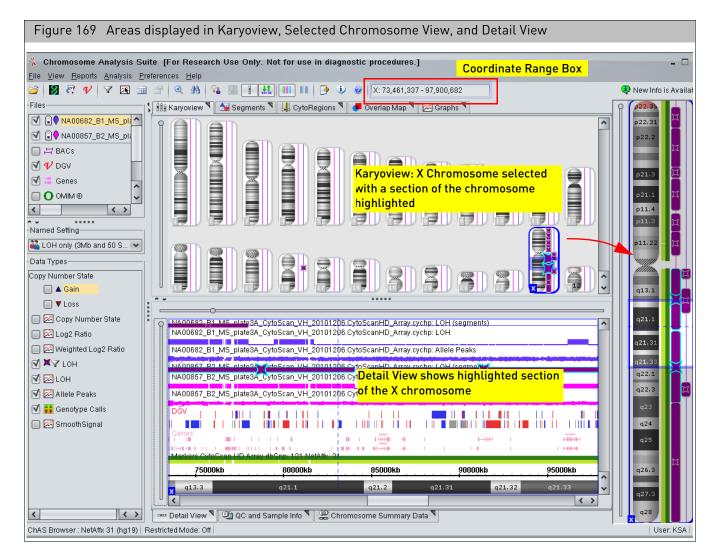
• Click a chromosome in the Karyoview.

The chromosome is displayed in the Selected Chromosome View and the Detail view.

#### Examining 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 169).



## Coordinate range box

The Coordinate Range box (Figure 170) is located in the ChAS main tool bar. 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.



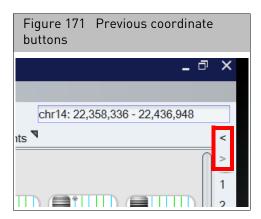
#### Going to a specific coordinate or coordinate range

- 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.

#### Returning to a previous location

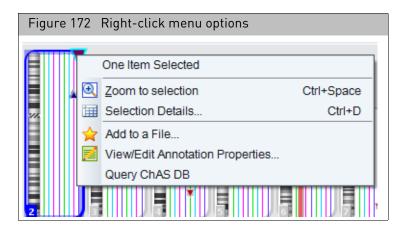
• To return to previous genomic coordinates, click on the <> buttons above the chromosome number field (far right), as shown in Figure 171.



#### Zooming to a selected item

There are several ways to zoom in on a feature.

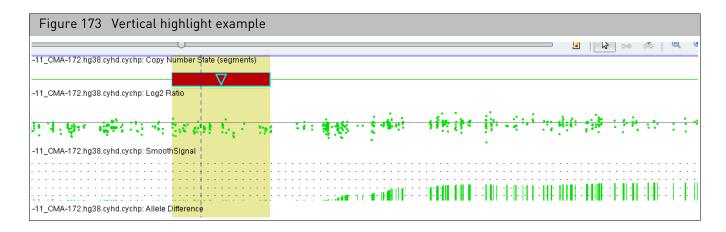
- Click a segment in the Segment table, CytoRegions table, or Overlap Map. Clicking on items auto-zooms to your configured zoom buffer (15% of item size is default), To edit the zoom buffer, see "Zoom buffer" on page 188.
- Double-click a feature in the Details View, Karyoview or Selected Chromosome View.
- Press **Ctrl+Space** to zoom all the way in to the selected item's start and end coordinates, with no buffer.
- In the Karyoview, Selected Chromosome View, or Details View, you can use the
   Zoom to selection option in the feature right-click menu (Figure 172) to go to the
   start and stop coordinates of the selected feature.



#### Setting a vertical highlight

Use this feature to set a vertical highlight on a selected segment(s) or annotation(s) to view its breakpoints across other data types and/or annotations. This highlighting option can be useful when aligning a segment/annotation of interest with other samples or annotations.

• Right-click on a selected segment(s) or annotation(s), then click **Set highlight region from selection** to apply a vertical yellow highlight through the Detail View, as shown in Figure 173.



### Edit configurations in the misc tab

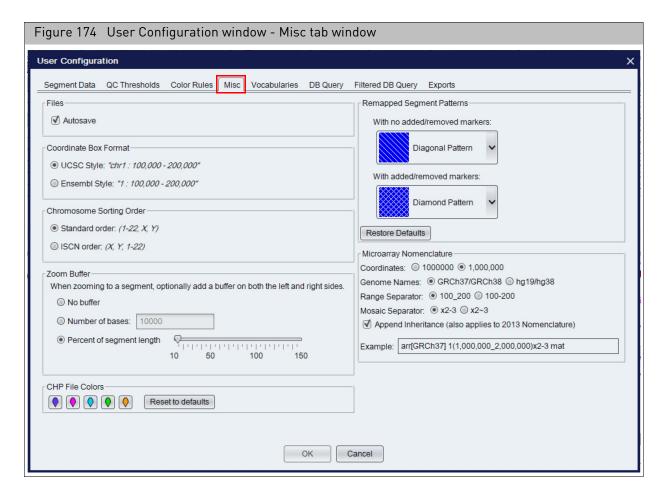
The Misc tab contains:

- Autosave
- Coordinate Box Format
- Zoom Buffer
- Chromosome Sorting Order
- CHP File Colors
- Remapping Segment patterns
- Microarray Nomenclature configuration

### Accessing the Misc tab

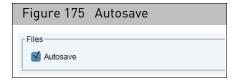
The User Configuration window appears. (Figure 174)

2. Click the **Misc** tab.



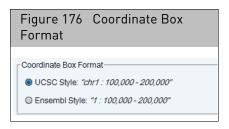
#### **Autosave**

• Click to check the **Autosave** check box to automatically save your files as they are edited. Uncheck to disable auto-save. (Figure 175).



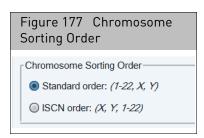
### Coordinated box format

 Click the appropriate radio button to choose the format of your displayed data. (Figure 176)



# Chromosome sorting order

Click the appropriate radio button to sort the segments in the Segment Table. (Figure 177)

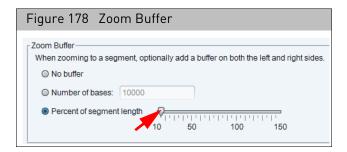


#### Zoom buffer

By default, the zoom percentage is set to 15% in new user profiles.

The Zoom Buffer feature offers 3 settings:

- No Buffer: Click this radio button to turn off the Zoom Buffer feature.
- **Number of bases**: Click this radio button, then manually enter the number of bases you want.
- **Percent of segment length**: Click, then drag the slider bar (Figure 178) to the zoom percentage you want.



#### CHP file colors



There are five preset CHP file colors assigned to your CHP data (Figure 179), but each default color, can be changed to a different color.

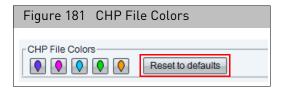
#### To change a CHP file color:

Click on the colored icon you want to change.
 A color wheel appears. (Figure 180)



- 2. Use the color wheel to locate the specific color you want or click on one of the several coloring options (including a website-safe color pallet).
- 3. Click OK.
- 4. Repeat steps 1-3 to change additional default colors.

At anytime, click **Reset to Default** to return the 5 CHP file colors back to their default colors. (Figure 181)



To return a single CHP file color back to its default color, click on the CHP file color, then click the Color Wheel's **Reset** button.

5. Click the User Configuration window's **OK** button to save your changes and exit.

**Note:** Samples that are currently loaded while a color change is made, may not reflect your new color scheme. To remedy this, close, then re-open ChAS to ensure your new color choices are reflected throughout all your samples.

## Remapped segment patterns

Choose how segments (remapped from an hg19 ChAS db to an hg38 ChAS db) are represented. See "Additional segment intersection information" on page 395 for information on default patterns.

- Click the With no added/removed markers: drop-down to select a pattern to represent segments in an hg38 ChAS db that were mapped from an hg19 ChAS DB, in which all the markers in the original segment mapped to the new genome version.
- 2. Click the **With added/removed markers:** drop-down to select a pattern to represent segments in an hg38 ChAS db that were mapped from an hg19 ChAS DB, in which one or more markers in the original segment no longer map to the new genome version.

#### Microarray nomenclature configuration

Define how the ISCN Microarray nomenclature will be represented in the segments table and exports. Any format from ISCN 2013, 2016 and 2020 can be configured

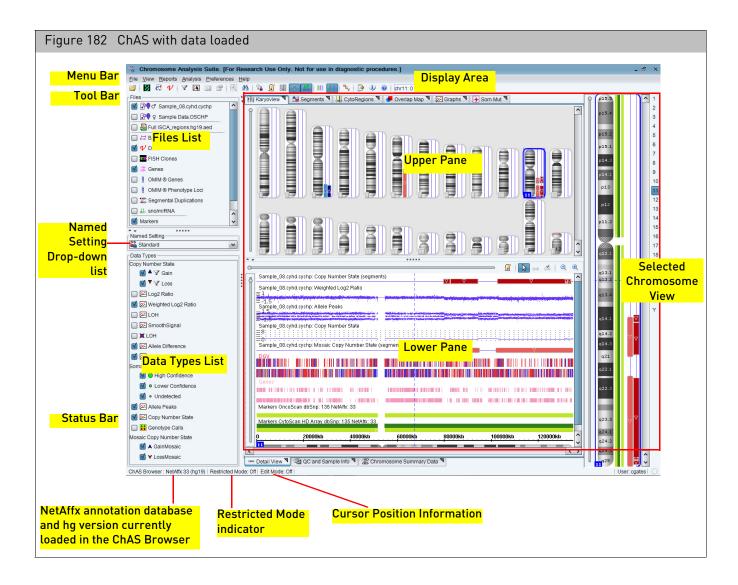
- Coordinates: choose the radio button to display genomic coordinates with or without commas.
- Genome names: choose the radio button to display genome versions using either GRCh or hg.
- Range separator: choose the radio button to display either a dash (-) or an underscore (\_) between genomic positions.
- Mosaic separator: choose the radio button to display either a dash (-) or a tilde (~) between copy number values on mosaic segments.
- (Optional) Select the check box to automatically have information in the Inheritance field appended to the Microarray nomenclature field.

### Controlling the display of data

ChAS (Figure 182) provides controls for:

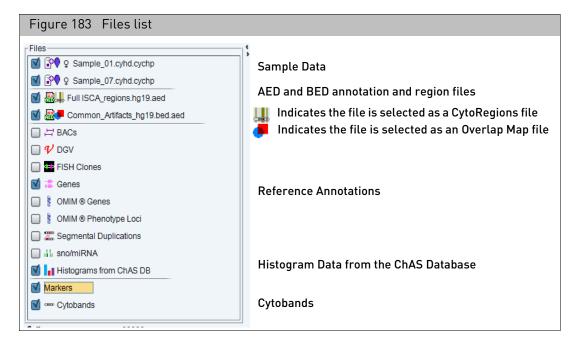
- "Selecting data for display" on page 192
- "Selecting data types for display" on page 194
- "Changing the grouping of samples and data types" on page 195
- "Selecting display schemes" on page 197
- "Expanding and contracting annotations" on page 197
- "Changing graph appearance" on page 198

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.



## Selecting data for display

The Files list (Figure 183) enables you to select sample data, region files, histograms, and reference annotations for display in the graphic view.



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.
- The appropriate gender symbol ♂ ♀ is displayed to the right of the colored nib. GenomeWide SNP 6.0 CNCHP file data may also contain a "?" if the gender was determined to be "unknown".
- 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 . VCF files are also loaded here.

- Histograms: Displays Gains, Loss, and LOH segments stored in ChAS DB. The
  display is loaded in the Detail View upon software start up or by selecting ChAS
  DB → Load Histograms (when connected to the ChAS DB).
- **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.

#### Selecting and deselecting files for display

1. Click in the check box 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.

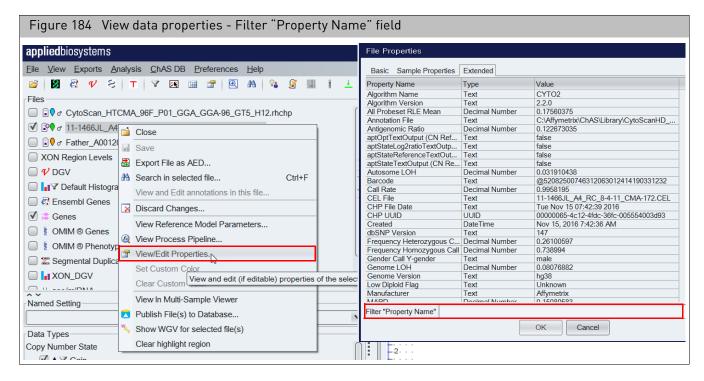
#### Changing the order of the Sample lanes or reference annotations

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

#### Viewing data properties

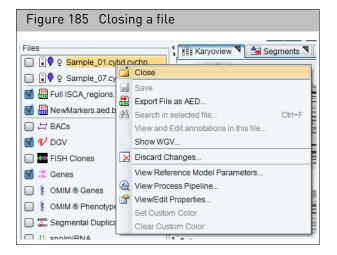
 Right-click a file and select View/Edit Properties on the shortcut menu. The Properties window appears.

(Optional) Use the Filter "Property Name" search text field (Figure 184) to quickly locate a property of interest.



#### Closing a file

- 1. Right-click on the file you want to close.
- 2. Select Close from the menu. (Figure 185)



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

# Selecting data types for display

The Data Types list (Figure 186) shows the data types that can be displayed in ChAS.

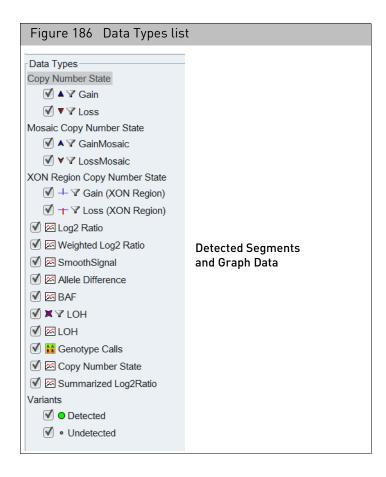


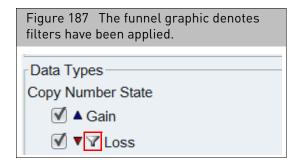
Figure 186 displays a list of data that can be displayed in the Karyoview, Selected Chromosome View, Detail View, and tables. The available data types may vary, depending upon the type of sample data available.

It enables 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 papears next to the segment type name, as shown in Figure 187.



The Graph data is displayed only in the Detail View.

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

• Click in the check box 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 "Named settings" on page 437).

#### 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.

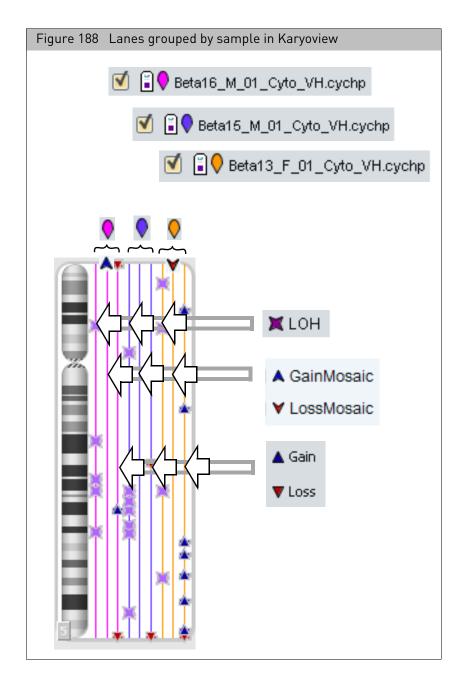
Each segment type is assigned to its own lane and has its own symbol.

#### Changing the grouping

From the View menu, select Group by Sample or Group by Type or in the tool bar, click the Group by Sample or Group by Type button.
 This enables you to perfom different types of comparisons between samples and segment types.

#### 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 (Figure 188) and in the Detail View.



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 and in the Detail View.

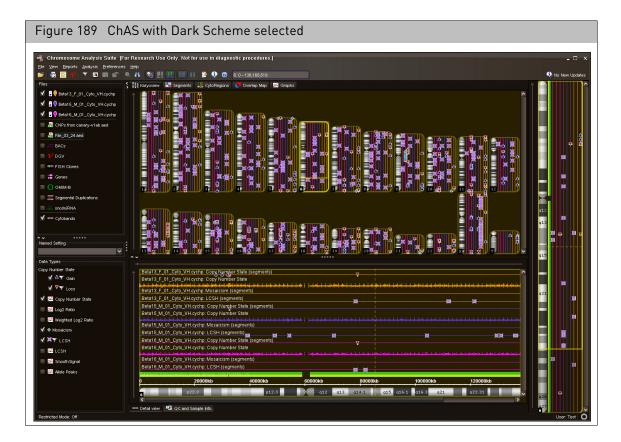
**Note:** You can change the order of samples and Data types in the views by clicking and dragging in the Files and Data Types list.

### Selecting display schemes

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

#### Selecting a dark scheme display

- 1. From the View menu, select **Use Dark Scheme**; or
- 2. Click the **Dark Scheme** button in the main tool bar. The Dark/Light scheme is used (Figure 189).



# Expanding and contracting annotations

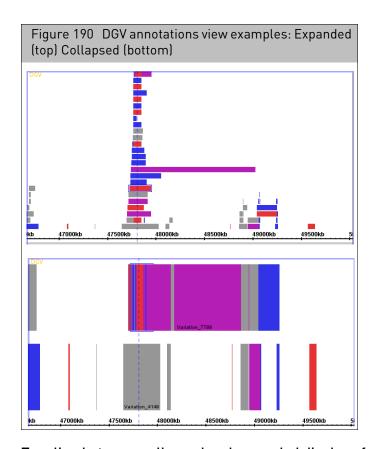
For tracks containing multiple rows of annotations, collapsing tracks consolidates all rows within a track into two rows. Any annotations after the first one will be placed in the second row (Figure 190).

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.

To expand or collapse just a single annotation track, right-click on the track name in the Files tree and choose Expand (or Collapse) Annotation track. The Annotation track check box must be checked in order to see this option in the right-click menu.

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



#### Toggling between collapsed and expanded display of annotations:

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

# Changing graph appearance

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

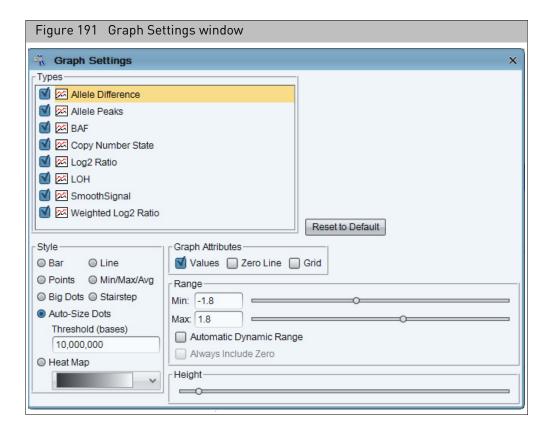
- "Selecting different graph styles" on page 199
- "Changing graph attributes" on page 203
- "Changing scale" on page 204

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

#### Opening the Graph Settings window

1. From the **View** Menu, select **Graph Settings** or click on the **Graph Settings** button in the Graphs Tab tool bar.

The Graph Settings window opens. (Figure 191)



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

#### Changing 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 styles" on page 199
  - "Changing graph attributes" on page 203
  - "Changing scale" on page 204

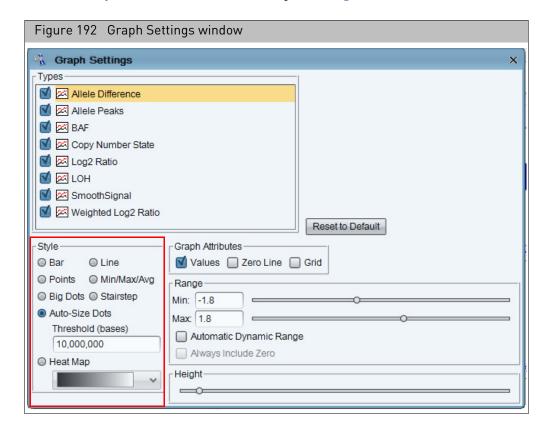
**Note:** Any changes you make to the values in the Graph Settings window will apply to all currently selected graph types.

#### 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.

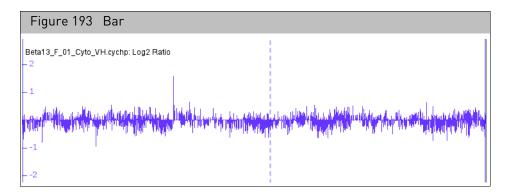
#### Changing the graph style

1. In the **Style** section choose one of the options. (Figure 192)

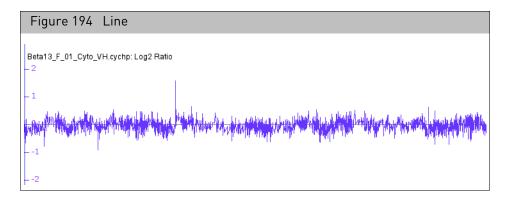


**Note:** Line, Min/Max/Avg, and Stairstep are not available for CytoScan XON arrays due to the differential coloring based on Level assignment.

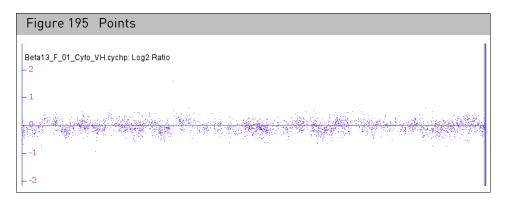
• **Bar** – Individual values are shown as vertical bars that are one base wide for position graphs. (Figure 193)



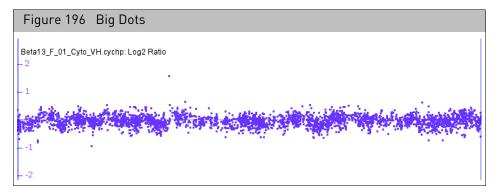
• **Line** – Subsequent values are linked with a line. Even if the input file was not sorted, the values will be connected in order along the genomic coordinate axis. (Figure 194)



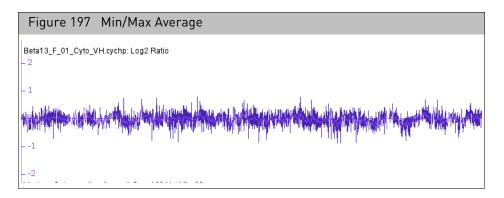
• **Points** – Shows a single dot for each data value. (Figure 195)



• **Big dots** – Shows a single big dot for each data value. (Figure 196)



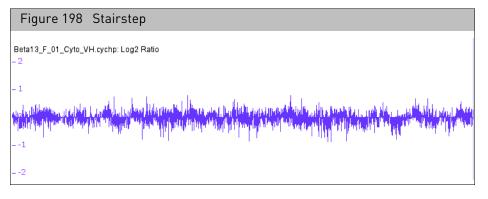
Min/Max/Avg – This style is especially useful for showing very densely populated graphs with data points for large numbers of positions. (Figure 197)
 Note: This data style is not available for CytoScan XON arrays due to regions annotated by Levels.



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.

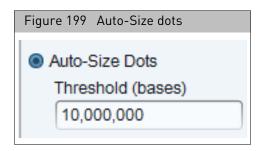
• **Stairstep** – Similar to the bar graph style, except that bar widths along the horizontal axis are stair-stepped. (Figure 198)

**Note:** This data style is not available for CytoScan XON arrays due to regions annotated by Levels.

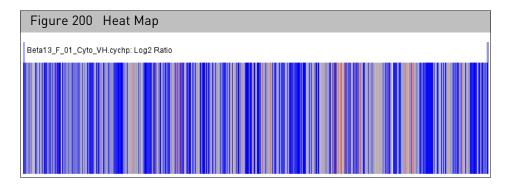


For example, if position 100 has a value of 50 and position 200 has a value of 75 and there are no values in between, then ChAS will draw a bar of height 50 that starts at position 100 and stops at position 200. Then, at position 200, ChAS will draw a new bar of height 75 that terminates at the next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

 Auto-Size Dots - Transition from Points to Big Dots when zooming in the Detail View. You can select the window size (in base pairs) in which the transition occurs. (Figure 199)



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 (Figure 200). 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.



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.

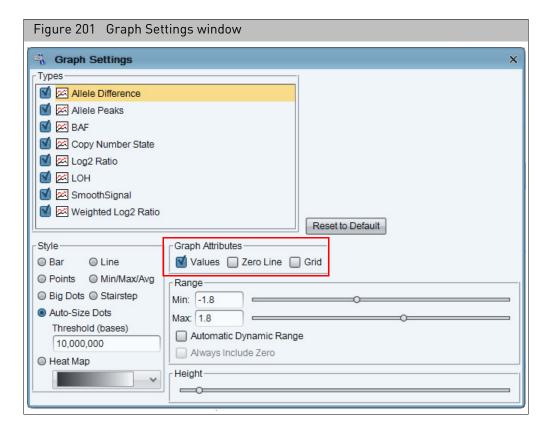
#### Changing graph attributes

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 201)



#### 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.



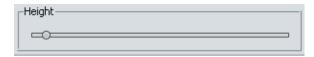
#### To set specific minimum and maximum values:

- 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.
  - Click the Automatic Dynamic Range check box to auto-set the Y-axis min/max range.
  - Click the Always include Zero check box to include a 0 point of reference on the Y axis.

**Note:** The algorithm detects CN state up to 4 for Genome-Wide Human SNP Array 6.0 CNCHP files and CytoScan Array CYCHP files. The TuScan algorithm detects CN State > 50 for OncoScan FFPE Assay OSCHP files. In the Graphs Settings, the Copy Number State range can be set to what is appropriate to the data.

#### Changing 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"
- "Right-click menu options" on page 208
- "Selection details table" on page 210
- "Linking to external websites" on page 215

#### 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
- Histograms
- Reference annotations

**Note:** You should expand the reference annotations before selecting one to avoid selecting multiple annotations. See "Expanding and contracting annotations" on page 197.

- Displayed Region files, including
  - Overlap Map 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, "Genomic position coordinates" on page 487).

You can learn more about the terms used in the pop ups in "Selection details table" on page 210.

#### Turning pop-ups on or off

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

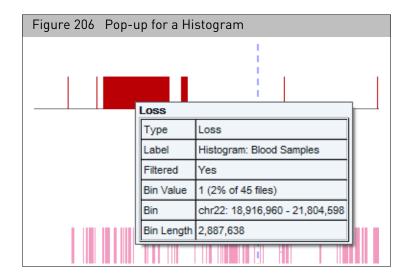
The information (Figure 202, Figure 203, Figure 205, and Figure 206) can include custom properties created by a user (see "Viewing and editing annotations" on page 298 for more information).

Figure 202 Pop-up for CN State Graph		
Sample_01.cyhd.cychp		
Array Type Cyt	oScanHD_Array	
Copy Number State		
Туре	Copy Number State	
NetAffx Version	33	
Array Type	CytoScan HD Array	
Nearest Coord	16,501,954	
Nearest Value	2	

Figure 203 Po	p-up for a Gain Segment	
Sample_01.cyhd.cychp		
Array Type CytoScanHD_Array		
Gain 11: 20,519,743 - 21,166,430		
Туре	Gain	
Location	11: 20,519,743 - 21,166,430	
Size (kbp)	647	
CN State	3.00	
Marker Count	958	
Median Log2Ratio	0.207	

Figure 204 Pop-u Segment	p for a XON Region Gain	
CytoScanXON_Sample.xnchp		
Array Type CytoScan XON		
Gain (XON Region) 1: 196,715,959 - 196,717,057		
Туре	Gain (XON Region)	
Location	1: 196,715,959 - 196,717,057	
Size (kbp)	1	
XON Region Level	1	
Marker Count	17	
Median Log2Ratio	0.147	
Summarized Log2Ratio	0.404	

	Figure 205 Pop-up for a Gene		
[	SHOX		
	Туре	Protein Coding Genes	
-	Location	X: 585,079 - 620,146	
	Size (kbp)	35	
	%HI	76.05	
	pLI	0.67	
1	CI	Not Available	
-[	Exon/Intron	Intron 1/5 →	



### 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 (Figure 207)
- Histograms (Figure 208)
- Reference annotations, including cytobands (Figure 209)

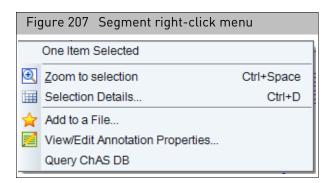
**Note:** You should expand the reference annotations before selecting one to avoid selecting multiple annotations. See "Expanding and contracting annotations" on page 197.

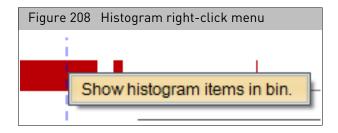
- Displayed Region files, including (Figure 210):
  - Overlap Regions
  - CytoRegions

**Note:** Not all options are available for the different feature types. Also, 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:

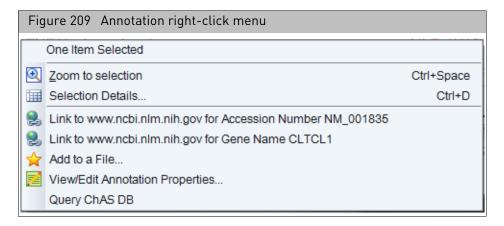
- Number of items selected (if more than one item is selected, the options available may differ, depending upon the type and number of items).
- **Zoom to selection:** See "Zooming to a selected item" on page 185.
- **Selection Details:** Opens the Selection Details box with information about the feature. See "Selection details table" on page 210.
- Add to a file: Add the selected segment, annotation, or region to a region file. See "Adding regions to an existing AED file" on page 293.
- View/Edit Annotation Properties: Displays the Annotations Properties window for the selected feature. You may or may not be able to edit the properties. See "Viewing and editing annotations" on page 298.
- Query ChAS DB: Displays the segments in the database that match the user defined Overlap and Coverage threshold settings. See "Setting up a ChAS DB query" on page 388.





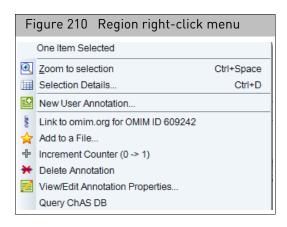
Special options for annotations include:

• Link to remote web site for more information. (Figure 209)



Custom functions for regions include (Figure 210):

- New User Annotation: Opens the New User Annotation window, which enables you to add user annotation to a region. For more information, see "New user annotations" on page 305.
- 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" on page 298.



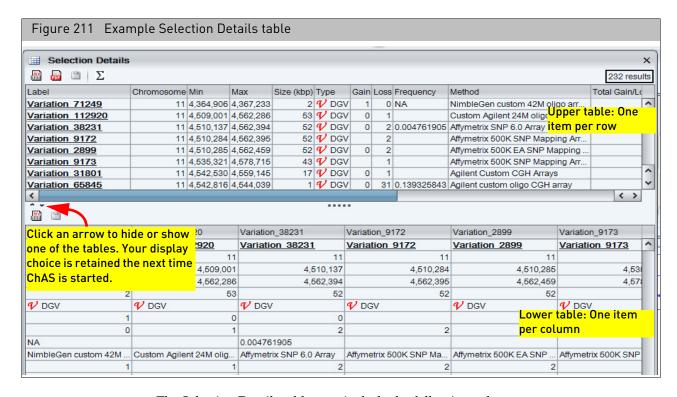
### Selection details table

The Selection Details table (Figure 211) 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.



The Selection Details table may include the following columns:

Column	Description
Common	
Label	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.

Column	Description
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, "Genomic position coordinates" on page 487).
Size (kbp)	Size of the item.
Туре	Type of segment (Gain, Loss, GainMosaic, LossMosaic LOH) or annotation.
Segments	
File	File the segment was detected in.
CN State	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 "LOH segments on X and Y chromosomes" on page 49 for more information.
Mean Marker Distance	Length of the segment in base pairs divided by the number of markers in the segment.
Interpretation	User-editable field for free-text interpretation on the segment
Call	User-editable field populated by a user-configurable drop list of Calls.
Inheritance	User-editable field populated by a user-configurable drop list of Inheritance.
Curation By	The current computer Operating System login ID and ChAS user profile name at the time that the Call or Interpretation field was last edited.
Curation Time	The time and date when the Call or Interpretation field was last edited.
Materially Modified Segment	Indication that segment was previously merged, deleted, or had its start or end boundary, type, or state altered by a ChAS user. (ChAS-based processes of Smoothing and Joining are not "Modifications", nor are making Calls or Interpretations, in this context).
Materially Modified By	The current computer Operating System login ID and ChAS user profile name at the time that the segment was last materially modified.
Materially Modified Time	The time and date when the Segment was last materially modified.
Max % Overlap	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".
Overlap Map Items (% of Segment overlapped)	Item(s) in the Overlap Map which overlap the segment, followed by the percentage by which the segment is overlapped by that Item.
CytoRegions	Names of the CytoRegions with which the segment shares coordinates.
Use in Report	Allows manual selection of Segments for export to a Segments Table PDF, DOCX, or Text rather than all segments in the table.
Marker Count	Number of markers in the segment.
Cytoband Start	Cytoband in which the segment begins.

Column	Description
Cytoband End	Cytoband in which the segment ends.
Genes	List of RefSeq genes from the Genes track that share coordinates with the segment. Identically named gene isoforms are NOT repeated.
Gene Count	A count of the gene names listed in the Genes column
DGV	List of DGV variations that share coordinates with the segment.
sno/miRNA	List of sno/miRNA features that share coordinates with the segment.
Old OMIM	The column formerly known as "OMIM" in the ChAS software when NetAffx Genomic Annotation Browser files version NA32.1 and earlier are loaded. This column's information is now out of date and has been superseded by the newer OMIM Genes and OMIM Phenotype Loci columns present in NA32.3 and above browser annotation files.
OMIM Genes	List of OMIM Genes that share coordinates with the segment.
OMIM Gene Count	A count of the OMIM Gene names listed in the OMIM Genes column.
OMIM Phenotype Loci	List of OMIM Phenotype Loci that share coordinates with the segment.
OMIM Region Phenotype Loci	List of OMIM Phenotype associated with a genomic region. The OMIM Morbidity information is displayed when using all three OMIM tracks. (OMIM Genes, OMIM Phenotype Loci and OMIM Region Phenotype Loci)
Segmental Duplications	List of Segmental Duplications that share coordinates with the segment.
Smoothed/Joined	Indication that segment was created by smoothing or joining two or more segments in the initial segment detection.
Segment Label	A label comprised of the segment's Type, State, and Filename.
Segment Name/ID	File-specific identifier assigned to the detected segment.
Start Marker	The array marker name which marks the beginning of the segment.
End Marker	The array marker name which marks the end of the segment.
Preceding Marker	The array marker just above the segment in the data track used as input for the segment.  Note: This column is only applicable to CNState Gain and Loss segments.
Preceding Marker Location	The coordinate location of the array marker just above the segment in the data track used as input for the segment.  Note: This column is only applicable to CNState Gain and Loss segments.
Following Marker	The array marker just below the segment in the data track used as input for the segment.  Note: This column is only applicable to CNState Gain and Loss segments.
Following Marker Location	The coordinate location of the array marker just below the segment in the data track used as input for the segment.  Note: This column is only applicable to CNState Gain and Loss segments.
Mean Log2 Ratio	The mean of all the Log2 Ratio values contained in the segment.
Mean Weighted Log2 Ratio	The mean of all the Weighted Log2 Ratio values contained in the segment.
Microarray Nomenclature	An ISCN-based description of the segment.

Column	Description
Sample UUID	Unique identifier for the CHP file.
Max % Coverage	The highest percentage by which a segment covers some item(s) in the Overlap Map.
Number of Overlap Map Items	Number of Overlap Map items which share genomic coordinates with the segment.
% of Overlaps Map Item covered by Segment	Overlap Map Item and the percentage by which it is covered by the segment.
Full Location	Chromosome Start and Stop in a user-friendly format for use in external databases.
Median log2	The median of all the Log2 Ratio values contained in the segment.
DB Count Both	Number of segments in the database meeting both the user defined thresholds of minimum Percent Overlap Count and Coverage Count.
DB Coverage Count	Number of segments in the database meeting the minimum Percent Coverage Count.
DB Overlap Count	Number of segments in the database meeting the minimum Percent Overlap Count.
XON Region Level	The annotation Level assigned to this region of the genome.
Summarized Log 2 Ratio	the median of the LR, after transformation to adjust for individual marker responsiveness.
Genes	
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, "Genomic position coordinates" on page 487).
Size (kbp)	Size of the item.
Туре	Type of segment (Gain, Loss, GainMosaic, LossMosaic LOH) or annotation.
Accession Number	Unique identifier assigned to the sequence in GenBank.
CDS Min	Minimum position of the coding sequence (BED-style coordinates).
CDS Max	Maximum position of the coding sequence (BED-style coordinates).
Strand	The sequence strand of the item.
Ensembl Genes	
CDS Min	Minimum position of the coding sequence (BED-style coordinates).
CDS Max	Maximum position of the coding sequence (BED-style coordinates).
Strand	The sequence strand of the item.

Column	Description
ОМІМ	
OMIM Gene Title	The title of the gene associated with the OMIM entry.
OMIM Gene Symbol List	A list of genes associated with the OMIM entry.
OMIM Disorder	Disorder associated with the OMIM entry.
OMIM Phenotype Key	Indicates how this phenotype was placed on the map.
OMIM Gene Symbol	Symbol of the gene based on gene title.
OMIM Phenotype LOCI	
OMIM Phenotype ID	Unique identifier to an OMIM phenotype.
OMIM Phenotype Map Key	Indicates how this phenotype was placed on the map.
OMIM Phenotype Locus Description	Describes the phenotype or disorder associated at the OMIM Phenotype Loci.
OMIM Region Phenotype I	Loci
OMIM ID	Unique identifier to an OMIM phenotype.
Cytoband	Location of the region.
Min	Start location of Genome.
Max	End location of Genome.
Gene Symbol	Symbol of the gene (based on gene title).
Gene Title	The title of the region associated with the OMIM entry.
Comment	Additional information for the entry from OMIM.
Gene-Phenotype Loci Data	Information of Gene-Phenotype relationships for the region.
Segmental Duplications	
Score	Score based on the raw BLAST alignment score. The score for segmental duplications is set to zero in NetAffx annotation 31 and higher.
FracMatch	The fraction of matching bases.
FracMatchIndel	The fraction of matching bases with Indels.
Strand	The sequence strand of the item.

**Note:** Thermo Fisher Scientific does not generate or verify the information for genes, FISH clones, Segmental Duplications, sno/miRNAs, DGV annotations, or OMIM data. 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" on page 298).

You can export data from the table using the standard table export tools (see "Exporting table data" on page 420).

You can perform multi-column sorts. See "Sorting by columns" on page 328.

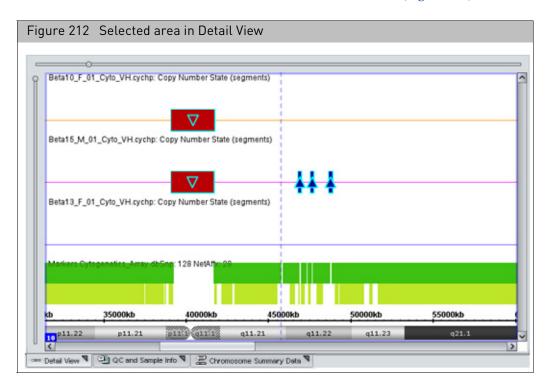
### Linking to external websites

You can view a selected area within the Detail View at one of the following public sites:

- 🌠 UCSC
- <equation-block> Ensembl
- V Toronto DGV
- 8 ClinVAR
- ClinGen
- DECIPHER
- T Load TaqMan Assays Track

#### To view the selected area at a public site:

1. In the Detail View, zoom and scroll to the area of interest. (Figure 212)



2. From the View menu, select **View Region at [site name]** or click the appropriate site's tool bar button.

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

### Linking to TaqMan copy number and genotyping assays

Viewing and ordering TaqMan assays for CN

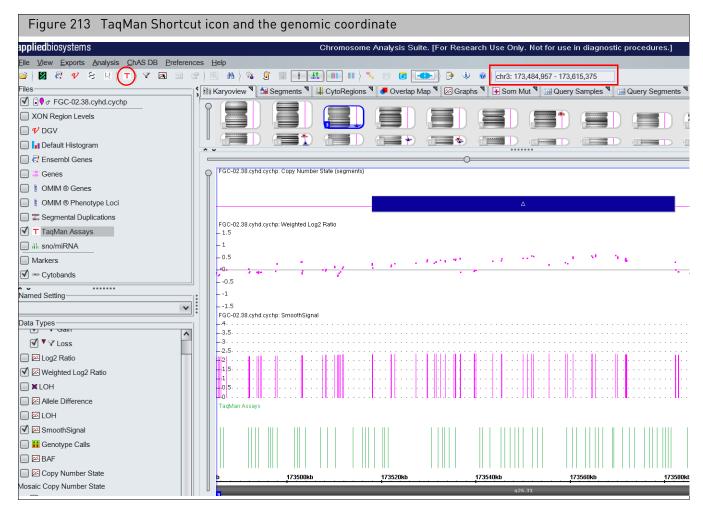
TaqMan assays can easily be accessed from within ChAS. These assays can be used for confirmation of copy number aberrations. TaqMan assays can only be ordered based on hg38 genome coordinates.

Do the following for the region(s) you would like to view and order TaqMan assays for Copy Number:

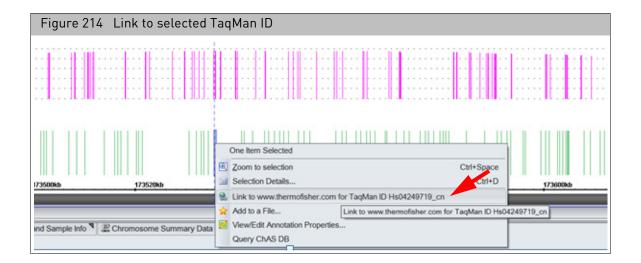
- 1. Locate the region containing the aberration in the Detail View.
- 2. Select the TaqMan shortcut in the tool bar (Figure 213) to view which TaqMan assays are available within the genomic coordinates populated in the text box at the top center of the browser.

The TaqMan assays load in a track in the Detail view.

**Note:** This track is only available for the genomic coordinates for the current query. To view TaqMan assays in another region of the genome, repeat steps 1 and 2.



3. Right-click on the TaqMan assays you want to order, then click **Order TaqMan assay** to link out to the website, as shown in Figure 214.



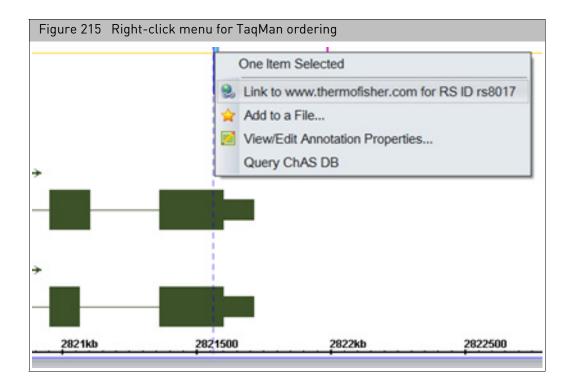
Viewing and Ordering TaqMan assays for genotyping TaqMan assays for genotyping can be ordered from VCF files that contain dbSNP IDs.

**Note:** The VCF file must contain an rsID for the SNP to directly access the TaqMan website for that SNP. Also, TaqMan assays can only be ordered based on hg38 genome coordinates.

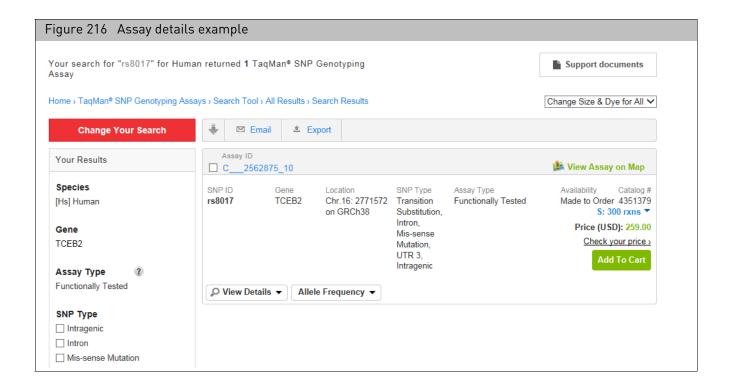
Do the following for the region(s) you want to view and order TaqMan assays for genotyping:

- 1. Load a VCF file clicking **File**  $\rightarrow$  **Open**.
- 2. Right-click on an SNP for which you would like to view and order a TaqMan assay.

A menu appears. (Figure 215)



Click on the provided "Link to" link.You will be directed to the TaqMan website for details about the assay.





## 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
- XON Level Assignment (XNCHP only)

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.

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 (genomewide). See "Using filters with CytoRegions" on page 276.

The Overlap Map filter is described in "Using the overlap map and filter" on page 280.

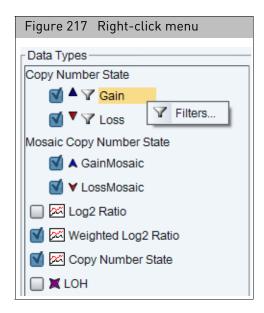
Filter settings are saved when a Named Setting is created and can be reapplied. See "User profiles and named settings" on page 436.

**IMPORTANT!** The Filters set in the browser are NOT linked to the filters for the MSV. The same filter settings should be set in both the ChAS browser and the MSV separately. The MSV does have a flag to indicate when filter settings do not match. For more information, see the RHAS User Guide.

### Applying segment parameter filters

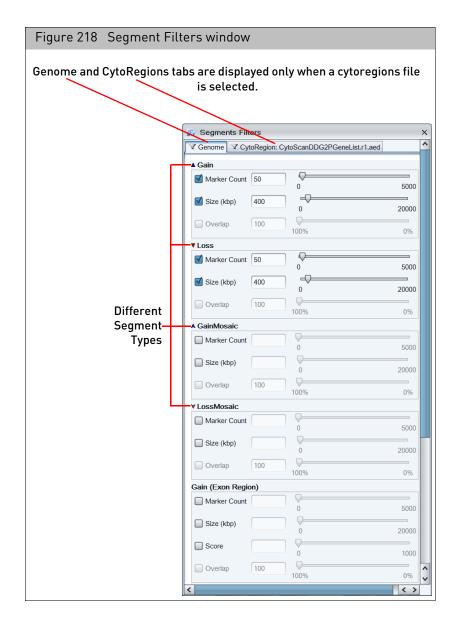
Opening the Segment Filters window

- 1. Select  $View \rightarrow Segment \ Filters$  on the menu bar
- 2. Click the **Segment Filters**  $| \gamma |$  tool bar button
- 3. Right-click on a segment type in the Data Types list, then select **Filters...** from the right-click menu. (Figure 217)



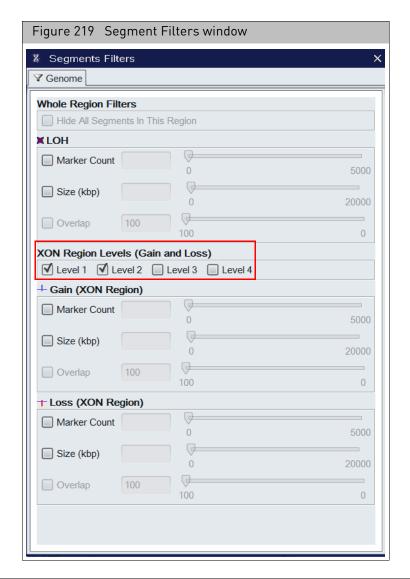
**Note:** If you use the right-click menu option, only the filter settings for the selected segment type are displayed.

The Segments Filters window opens. (Figure 218)



**Note:** The Overlap Map filtering parameter is set using the same window. The Overlap Map function is described in "Using the overlap map and filter" on page 280.

 For XON, check the Level check box(es) (as shown in Figure 219 on page 222) to reveal any XON region segment calls in regions assigned to the Level(s) you selected.

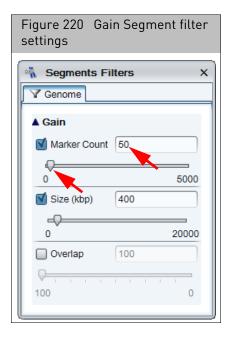


Segment Filter Option	Function
Hide All Segments in this Region	Hides all the segments. This is particularly useful when using a CytoRegions file for CytoScan XON arrays. Check the Hide All Segments in this Region check box on the Genome Tab so that ONLY segments overlapping CytoRegions will be shown. <b>Note:</b> This option is only available with a CytoRegions file is assigned.
Marker Count	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. 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.

Segment Filter Option	Function
Size	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.
XON Segment Level	Based on the Level Assignment to the region in the genome.  Level 1: Medical Research exome and cancer  Level 2: ClinVar genes not covered in Level 1  Level 3: Other OMIM genes  Level 4: Opportunistic regions from Refseq/UCSC/Enseble/LOVD.  The XON Segment Filters can be used to narrow down the number of XON segments to review based on their annotation level assignment. Those regions assigned as Level 1 contain genes/ regions that have been identified as part of the Medical Research Exome along with regions associated with cancer. By selecting only Level 1 in the filter settings, only XON segment calls in regions assigned as Level 1 will be displayed. XON segments for all other Levels will be hidden from view as well as hidden in the Segments Table. To expose XON segment calls in other regions simply check the box for those Levels in the Filters Windows. For more details on CytoScan XON analysis workflow recommendations, Appendix H, "Recommended CytoScan XON array workflows" on page 510.

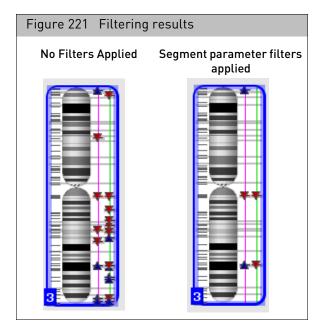
## Using segment parameter filters

1. Click the check box next to the parameters you want to use as filters. (Figure 220)



2. Use the slider to set the value for the parameter or enter a value in the provided text field. (Figure 220) **Note:** As you move the slider from left to right, more segments are removed.

Your filtered results are displayed instantly in all tables and graphs, as shown in Figure 221.



For information about using the Overlap setting, see "Using the overlap map and filter" on page 280.

For information on using different filtering settings in CytoRegions, see "Using filters with CytoRegions" on page 276.



## Segment modification

#### This chapter includes:

- "Editing segment data overview"
- "Using edit mode" on page 226
- "Types of segment editing" on page 227
- "Merging all segments types" on page 228
- "Deleting all segments types" on page 234
- "Editing the start/end Coordinates of all segment types" on page 237
- "De Novo segment drawing" on page 239
- "Changing all copy number segment types" on page 242
- "Promoting mosaic segments" on page 244
- "Modified segments in the segments table" on page 245
- "Removing all edits made to a sample" on page 246

### Editing segment data overview

- ChAS enables you to edit data segments; merge, delete, draw de novo, adjust boundaries, and change the Copy Number State of CHP data segments, or undo your changes.
- The instant a file is modified by editing of one or more segments, having a
  segment interpreted, or called, ChAS auto-generates a "sidecar" or CHP Change
  Archive (CHPCAR) file which saves these modifications. ChAS then uses this
  CHPCAR file for further user-defined edits and modifications, while your
  original (native) xxCHP file (CYCHP, CNCHP, or OSCHP) is safeguarded
  (remains un-touched).
- The smoothing, joining and XON merging settings for each CHP file are locked at the time of CHPCAR file creation, while unmodified CHP files without CHPCAR files will still respond to array type-specific changes in smoothing and joining settings.
- In order for the changes stored in the CHPCAR file to show up in the CHP file
  data displayed in the ChAS Browser, the CHP file and CHPCAR file must be
  stored in the same directory/folder (the xxCHP file is still the only file that is
  required to be "loaded" into the browser, the CHPCAR file's edits, calls and
  interpretations will load when the CHP file is loaded).
- CHPCAR files are named using the entire CHP file's name, and contain the extension: \*.chpcar

- Changing the name of either file and not the other to match will disrupt the ability for the files to be recognized as associated with each other in the ChAS Browser.
- CHP files which have associated CHPCAR files detected (and in use), display as a special "CHP" icon in the File → Open dialog window, and in the Files tree of the Browser.
- Please move the CHP and CHPCAR files together when moving or archiving data.
- Two people using CHAS on different systems should NOT attempt to edit a CHP file at the same time.
- Modification made to segments in the ChAS browser will also be updated in the MSV.
- Files downloaded from ChAS DB can only use the delete segment option for editing segments.

**Note:** Segment modification does not apply to ReproSeq Aneuploidy data.

### Using edit mode

**IMPORTANT!** Make sure the Edit Mode feature is turned on BEFORE you start editing segments. Segments can not be published to ChAS DB while in Edit Mode. Edit Mode must be turned off to enable the Publish function.

**Edit Mode** is accessible in three places:

1. The Browser's icon row (top).



2. The Detail View's icon row (top right)

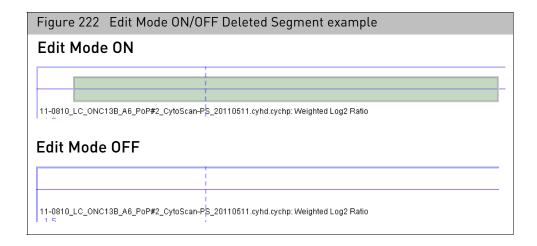


3. Click View  $\rightarrow$  Edit Mode

By default, Edit Mode is OFF. Click [1] (located on the Browser's top icon row or above the Detail View) to turn Edit Mode ON.

- Click 1 to turn Edit Mode OFF and remove all visual indications of your segment changes.
  - When Edit Mode is ON, deleted segments are visible, and edited segments appear distinct from non-edited segments.
  - When Edit Mode is ON, a track on a dotted axis line will appear showing the original calls made by the software for comparison with the manual modifications on the segment track.
  - When Edit Mode is OFF, deleted segments are invisible, and edited segments look identical to non-edited segments.

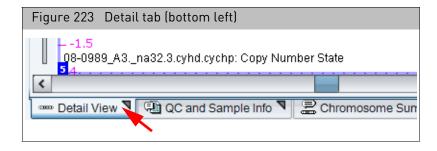
**IMPORTANT!** Turn Edit Mode OFF, before exporting a report of your data. Also, Edit Mode must be OFF, before publishing to the database.



### Types of segment editing

- "Tracking original calls" on page 228
- "Merging all segments types" on page 228
- "Deleting all segments types" on page 234
- "Editing the start/end Coordinates of all segment types" on page 237
- "De Novo segment drawing" on page 239
- "Changing all copy number segment types" on page 242
- "Promoting mosaic segments" on page 244
- "Editing the Microarray Nomenclature (ISCN 2013) and Microarray Nomenclature fields" on page 247

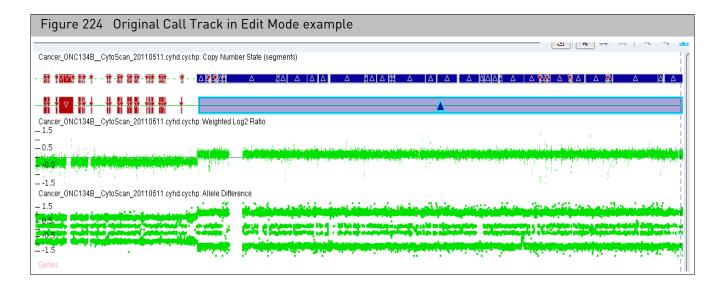
**Note:** Before you start any segment editing, make sure the Detail View tab is selected, as shown in Figure 223.



### Tracking original calls

Original calls can be tracked to view original segment calls made by the software.

- This track is only visible when Edit Mode is ON.
- This track is only for visualization in the Detail View and is not populated elsewhere in the software.
- The Original Calls track will disappear when the Edit Mode is OFF.



### Merging all segments types

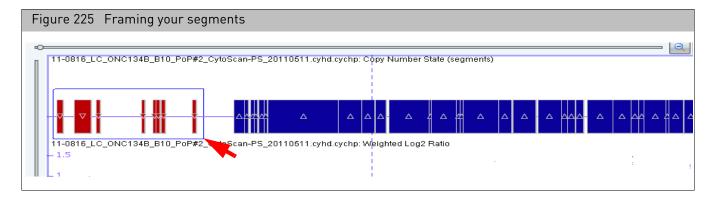
There are two ways to merge segments:

- "Merging segment groups"
- "Segment to segment merge" on page 231

## Merging segment groups

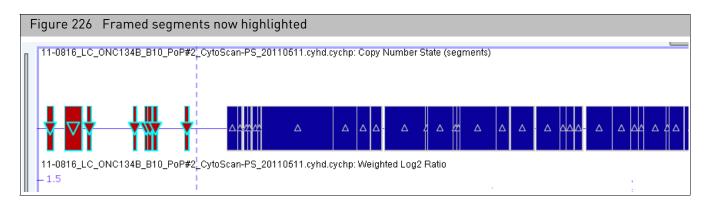
**Note:** Merging segment groups together, cancels out any previously assigned Calls. However, un-doing the group of merged segments (page 233) reinstates their original Calls.

- 1. Click **File**  $\rightarrow$  **Open**.
  - Your Sample File data folder window appears.
- 2. Click to select the file you want to edit, then click **Open**. The file appears in the ChAS browser's Detail view.
- 3. Left-click, hold, then move the mouse to frame the segments you want to merge together. (Figure 225)

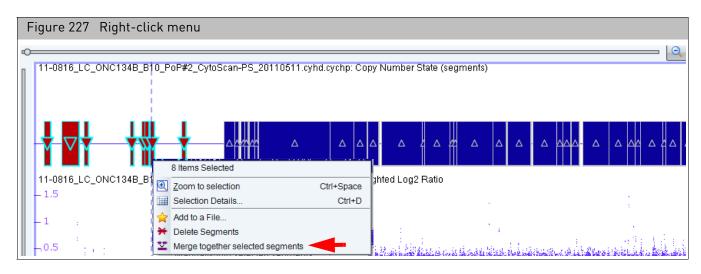


4. Release the mouse button.

Your selected segments, including their Loss symbols are now highlighted in **blue**, as shown in the example below. (Figure 226)



5. Right-click on a highlighted segment.
The following menu appears. (Figure 227)

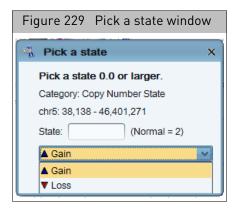


6. Click Merge together selected segments.

A message regarding your planned merge may appear. (Figure 228)



Click **OK** to acknowledge it.The Pick a state window appears. (Figure 229)



8. In the State field, enter a Copy Number, then use the drop-down menu to select a Type (Gain/Loss) for the new segment. (Figure 229)

**Note:** For CytoScan XON arrays (XNCHP files), it is not required to enter a copy number state value.

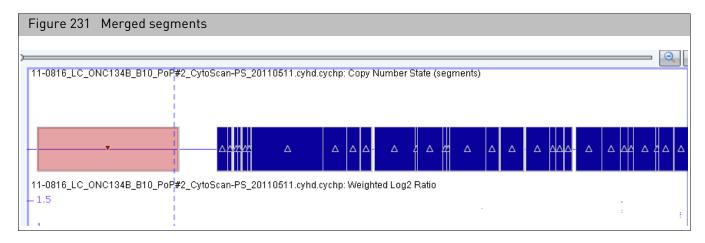
9. Click OK.

In cases where the segment to be merged into a group <u>contains</u> a **Call or Interpretation**, the following message appears: (Figure 230)



10. Acknowledge the message, then click **OK**.

11. Your selected segments are now merged together and appear as shown. (Figure 238)

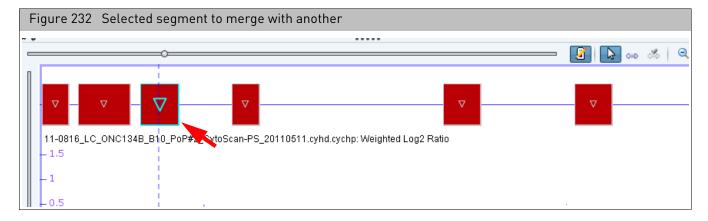


### Segment to segment merge

**IMPORTANT!** When merging segments during the editing of the start or end of one particular segment, only the segment whose start or end you are editing has the option of having its Call or Interpretation saved (or not).

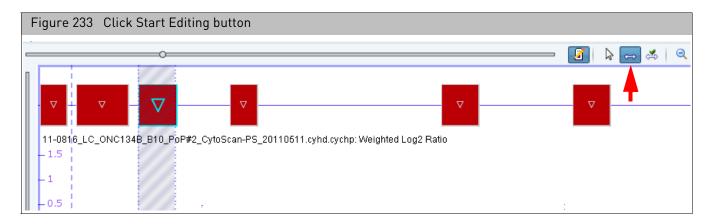
Segments which are being engulfed by the edit start/end procedure being performed will not have the option of having their Call or Interpretations placed on the resulting segment. However, un-doing the edit resulting in this type of merging (page 233) will reinstate previous Call and Interpretation information for all the original segments involved.

- Click File → Open.
   Your Sample File data folder window appears.
- 2. Click to select the file you want to edit, then click **Open**. The file appears in the ChAS browser's Detail view.
- 3. Click to select the segment you want to merge with another. (Figure 232)

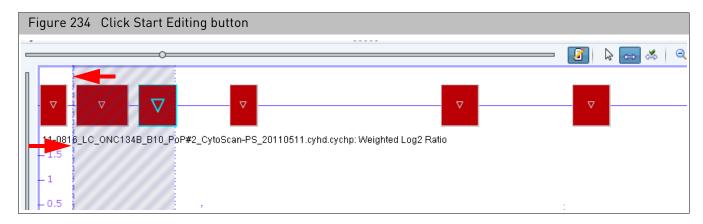


4. Click the **Start Editing** button.

5. A shaded area over your selected segment appears. (Figure 233)

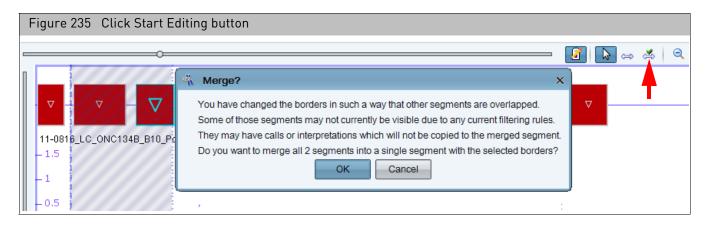


6. Click, then drag the shaded area over the segment you want to merge it with. Make sure you overlap your target segment just slightly, as shown. (Figure 234)



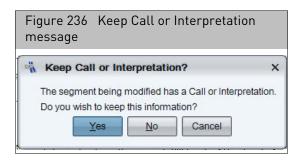
7. Click the **Start/End** button.

The following message appears: Figure 235



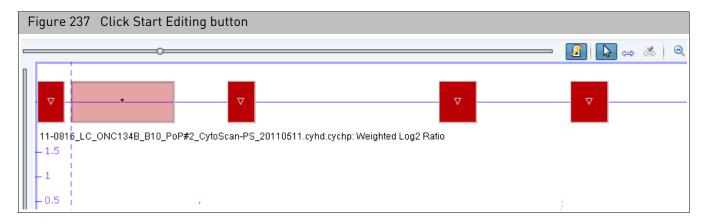
8. Acknowledge the message, then click OK.

If the segment whose Start or End you are editing <u>contains</u> Call or Interpretation information, the following message appears: Figure 236

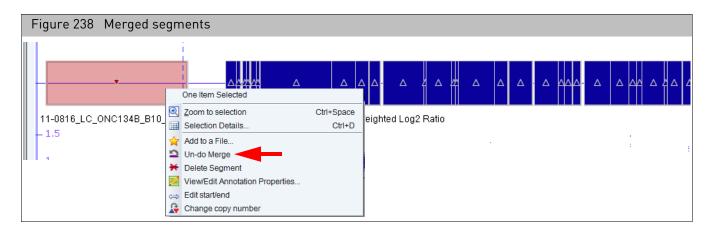


9. Click the appropriate button.

Your two segments are now merged together. While Edit Mode is **ON**, the segment appears in a distinct color. (Figure 237). (When Edit Mode is **OFF**, the edited segments look identical to non-edited segments.)

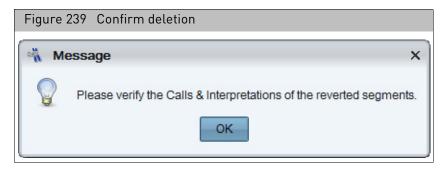


#### **Un-Merging Merged**



#### Segments

1. Right-click on the newly merged segments (Figure 238), then click **Un-do Merge**. The following message appears. (Figure 239)

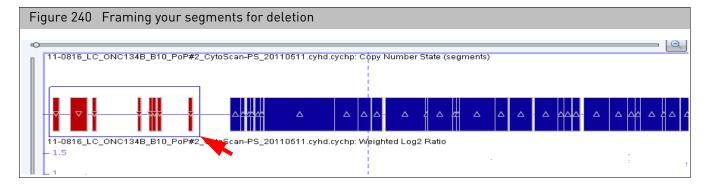


2. Acknowledge the message, then click **OK**.

Your previously merged segments return to their original states, including any Calls or Interpretations made on the original segments (prior to their merging).

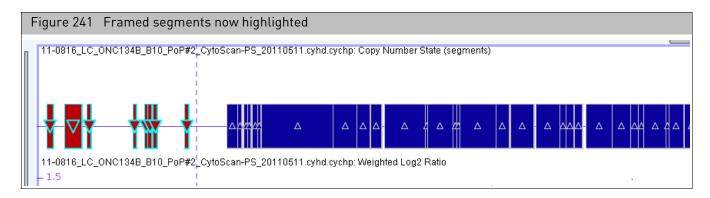
### Deleting all segments types

- Click File → Open.
   Your Sample File data folder window appears.
- 2. Click to select the file you want to edit, then click **Open**. The file appears in the ChAS browser.
- 3. Left-click, hold, then move the mouse to frame the segments you want to delete. (Figure 240)

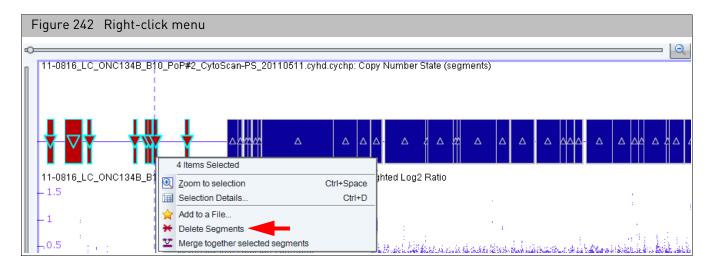


4. Release the mouse button.

Your selected segments, including their Loss symbols are now highlighted, as shown in the example below. (Figure 241)

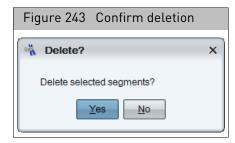


5. Right-click on a highlighted segment.
The following menu appears. (Figure 242)



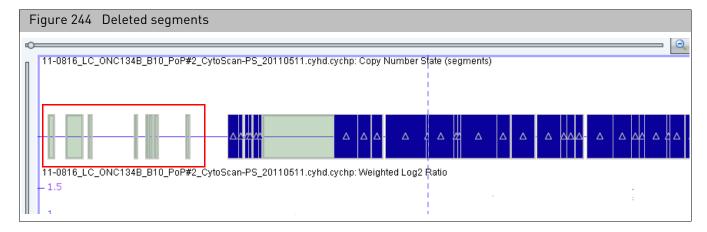
6. Click **Delete Segments**.

A message regarding your planned deletion appears. (Figure 243)



7. Click **OK** to acknowledge it.

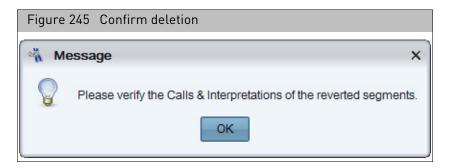
Your selected segments are deleted and graphically represented as "ghosts" while Edit mode is **ON**. (Figure 244). When Edit Mode is OFF, deleted segments are invisible and NOT displayed in any view or table.



**Note:** Even though your selected segments are deleted, ChAS preserves their data for future reference. See "Right-click menu options" on page 208 for information on how to view and edit segment properties and view segment details.

## Un-deleting a deleted segment(s)

Right-click on the deleted segment(s), then click Un-do Delete.
 The following message appears. (Figure 245)



2. Acknowledge the message, then click **OK**.

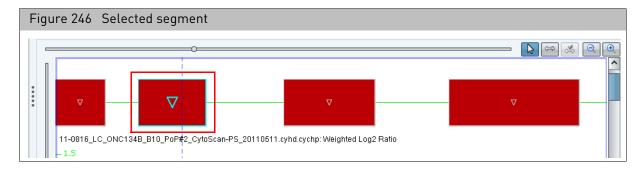
Your previously deleted segment(s) return to their original states, including any Calls or Interpretations made on the original segments (prior to their deletion).

### Editing the start/end Coordinates of all segment types

IMPORTANT! Only the boundaries of 1 segment, can be adjusted/modified at one time.

**Note:** Segment start or end boundaries can be moved left and/or right. The Adjusting/ Modifying Segment Boundaries example that follows, shows how to move a segment end boundary farther to the right.

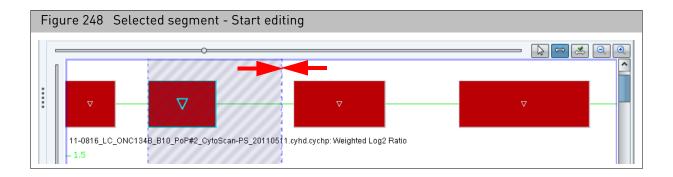
- 1. Using the zoom tools and scroll bars (if needed), identify the segment whose start or end (or both) you want to modify
- 2. Single-click on the segment to highlight it. (Figure 246)



- Click the Start editing... button.
   The following appears: (Figure 247 on page 237)
- 4. Place the mouse cursor on the very-right edge of the current segment boundary. (Figure 247)

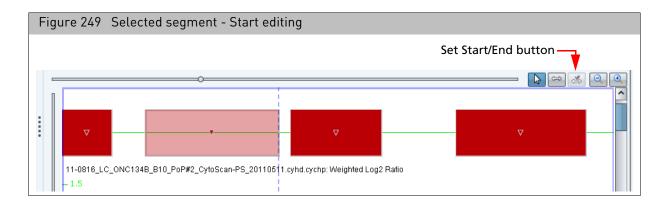


5. Click, then drag the segment's right-edge boundary to the right, then stop at an appropriate point. (Figure 248) You will ONLY be allowed to set the Start or End position to match the position of a marker probeset.

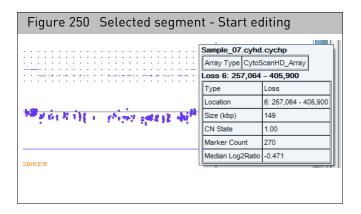


6. Click the **Set start/end**... button. (Figure 249)

The newly adjusted/modified segment boundary appears as shown. (Figure 249)



7. ChAS auto-calculates various properties of the newly edited segment. Mouse over the segment to view its new details. (Figure 250)



Un-doing the edited start/end coordinates of a segment

Right-click on the previously adjusted/modified segment, then click Un-do Edit.
 The segment returns to its previous state, including, previous boundaries, calls, and interpretations.

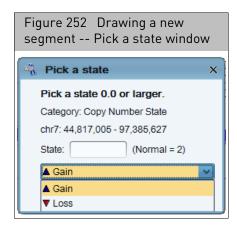
### De Novo segment drawing

- 1. Within an existing segment track, locate an empty space where you want to draw in a brand new segment.
- Right-click inside this empty space.
   A Create a Copy Number State segment around XX,XXX,XXX ribbon appears (Figure 251)



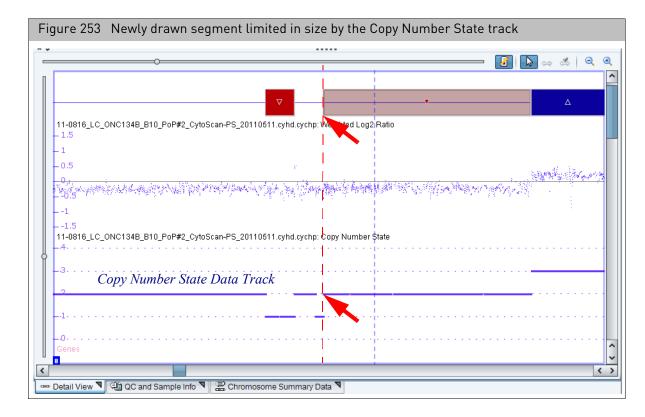
3. Click on the ribbon, as shown in Figure 251.

The **Pick a state** window appears. (Figure 252)



- 4. Enter a **Copy Number** and **Type** for the new segment.
- 5. Click OK.

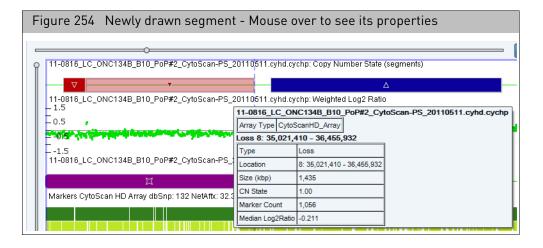
Your newly drawn segment appears. (Figure 253)



**IMPORTANT!** Newly drawn segments will stop when they encounter another segment, whether or not that segment is currently drawn or whether it is currently filtered out; see Figure 253 for an example in the CN State data track which is used to draw Copy Number Segments of Gain and Loss.

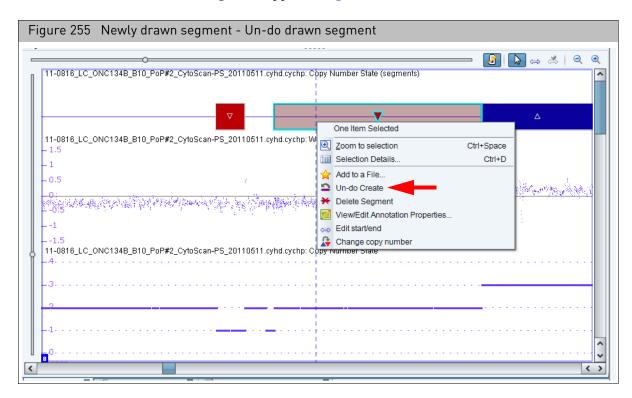
New segments will also stop when they encounter the last appropriate marker on the chromosome. Because of this, new segments drawn will vary in their initial size.

6. Move the mouse cursor over the newly drawn segment to reveal its properties. (Figure 254)



## Un-doing a segment De Novo drawing

1. Right-click on the newly drawn segment. The following menu appears. (Figure 255)



#### 2. Click Un-do Create.

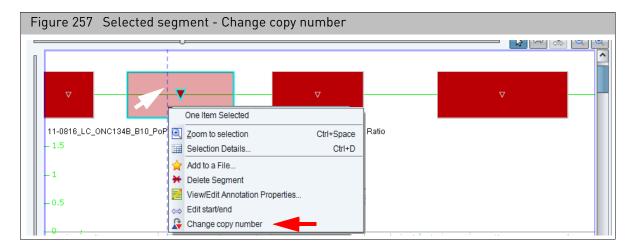
The previously drawn segment is removed, along with any Calls or Interpretations it had been given. (Figure 256)



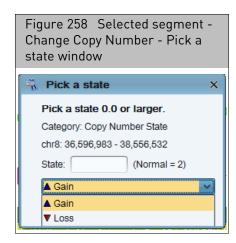
### Changing all copy number segment types

**Note:** For CytoScan XON arrays only, you are not required to assign a copy number to the segment.

- 1. Right-click on the segment that contains the copy number you want to change
- 2. Click Change copy number. (Figure 257)



The **Pick a state** window appears. (Figure 258)

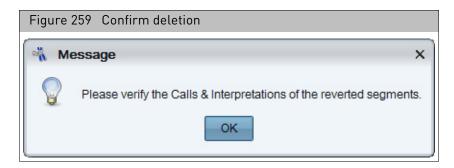


- 3. Enter a new Copy Number State value and Type (Gain/Loss). **Note:** For CytoScan XON arrays (XNCHP files), it is not required to enter a copy number state value.
- 4. Click OK.

The segment now reflects your revised Copy Number State and Type.

## Un-doing a copy number change

- 1. Right-click on the segment you performed a Copy Number State change on, then click **Un-do Edit**.
- 2. The following message appears. (Figure 259)

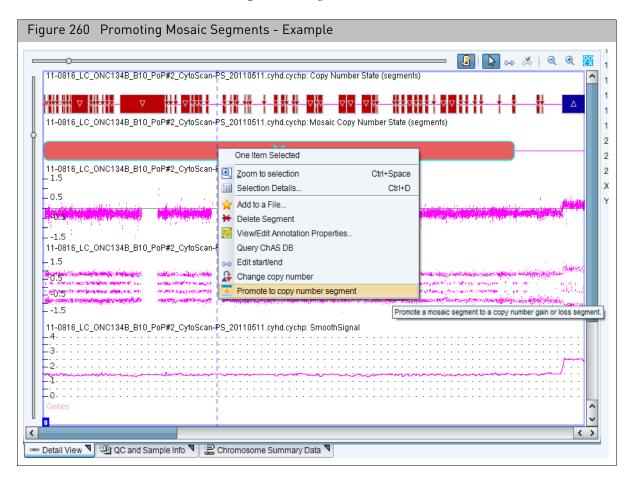


Acknowledge the message, then click OK.
 The segment's original Copy State Number is reinstated.

### Promoting mosaic segments

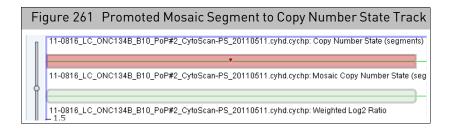
#### IMPORTANT! Segments in the Mosaic Segment Track are NOT uploaded to the database.

To capture the information for a mosaic segment in the database, that segment must be "promoted" to the Copy Number State track. This is done to reduce redundancy in those regions in which segments were called by both the copy number algorithm and the mosaic detection algorithm. (Figure 260)



Promoted mosaic segments maintain their non-integer copy number state, marker count, median log2 ratio, genome coordinates and size when they are promoted to the copy number state track. (Figure 261)

The mosaic gain segments will have the same blue/red used for integer copy number Gains/Losses, however, they maintain their non-integer copy number state to indicate they are a mosaic.



### Modified segments in the segments table

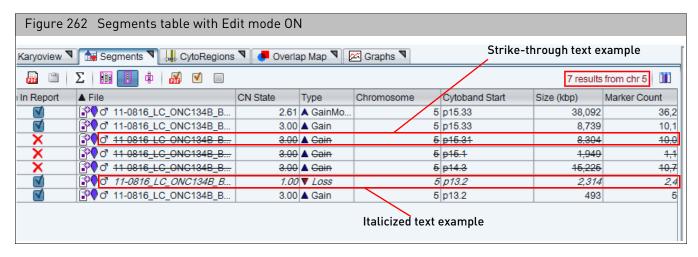
**IMPORTANT!** When exporting a Segments table to text (TXT), please note that deleted segments will be part of the export when Edit Mode is On, and will NOT be part of the export when Edit Mode is Off.

In PDF reports, deleted segments are never shown in graphical views, nor are they listed in the Segments Table, as <u>Edit Mode is required to be OFF to generate a PDF report</u>.

### Editing mode on

When the Edit Mode is ON \_\_\_\_\_, the Modified segments appear differently within the Segments table, as shown in Figure 262.

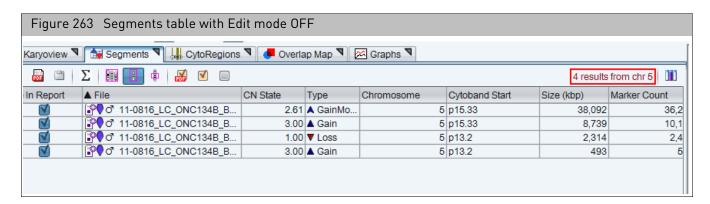
- Deleted Segments are represented with a red X and a strike-through line. Deleted segments do NOT show up in PDF reports because PDF reports cannot be created while Edit Mode is ON.
- Materially Modified Segments (including segments that have been merged, boundaries edited, or had their Copy Number States changed) are represented with *italicized* text.



#### Editing mode off

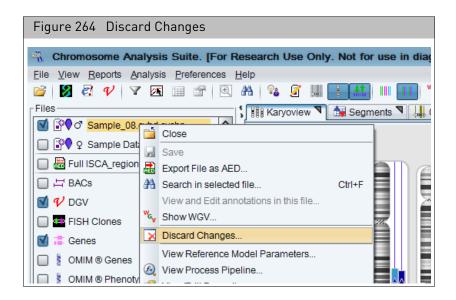
Turning the Edit Mode Grant OFF, displays the Segments table as follows: (Figure 263)

Note that the three deleted segment rows which were in strike-through text are removed from the table when Edit Mode is OFF. Also note, that the rows which indicated Modified segments are no longer italicized when Edit Mode is OFF.



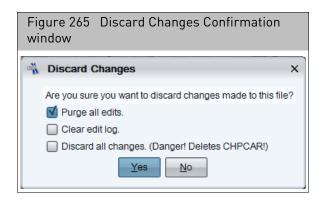
### Removing all edits made to a sample

1. Right-click on the File name in the Files list, then select **Discard Changes**. (Figure 264)



The Discard Changes window appears and displays the following options:

- Purge All Edits: Reverts all edits to segments and removes all calls and interpretations.
  - **Note:** The log of the edits that have been performed on the sample can still be viewed.
- Clear Edit Log: Purges all edits and clear the log of any edits performed on the sample.
- Discard all changes: Purges all edits, clears the edit log, and deletes the CHCAR file.
- 2. Check the appropriate box, then click **Yes**. (Figure 265)



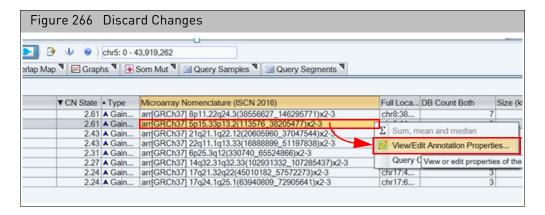
## Editing the Microarray Nomenclature (ISCN 2013) and Microarray Nomenclature fields

The View/Edit Annotation Properties Curation tab enables you to update the copy number nomenclature in the Microarray Nomenclature (ISCN 2013) and Microarray Nomenclature fields.

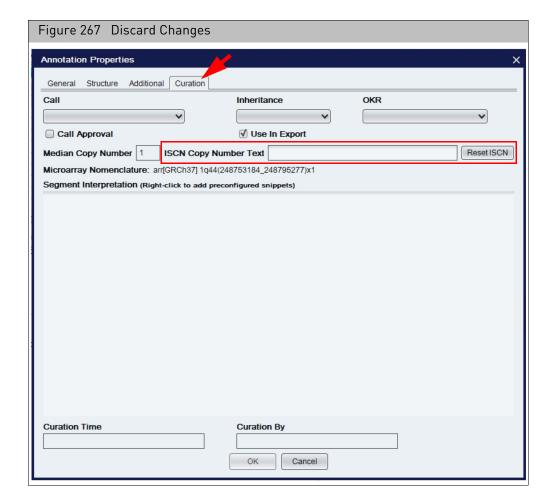
**Example**: For Mosaic segments, the default name will show a range of copy number such as x2-3. If you have the percent mosaicism information, this part of the Microarray Nomenclature can be updated to reflect the percent mosaicism by typing x3[0.6]. The field will be updated as follows:

Default name: arr[GRCh37] 5p15.33p13.2(113576\_38205477)x2-3 Updated name: arr[GRCh37] 5p15.33p13.2(113576\_38205477)x3[0.6]

1. Right-click on the Microarray Nomenclature cell for the segment you would like to update, then click **View/Edit Annotation Properties**, as shown in Figure 266.



The Annotation Properties window appears. (Figure 267)



- 2. Click the **Curation** tab.
- At the ISCN Copy Number Text field (Figure 267), type in the copy number nomenclature text you want, then click OK. Note: This field can be edited for LOH segments to designate copy number in the Microarray Nomenclature field for LOH segments as well.

To reset the text to the original default format, click the **Reset ISCN** button.

**Note:** To automatically add the Inheritance field to the Microarray Nomenclature ISCN 2016 field, please see "Adding or removing inheritance calls" on page 260.



## Sample and segment annotations

#### This chapter includes:

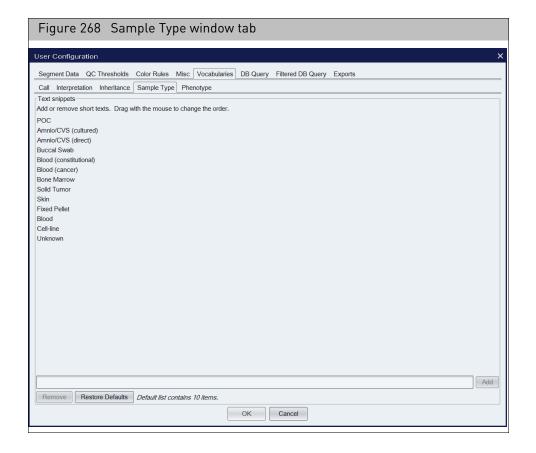
- "Sample annotations"
- "Segment annotations" on page 253
- "Tracking and reviewing the log file" on page 263

### Sample annotations

Sample file level annotations such as Sample-type, Phenotype, and Sample Interpretation can be added to each sample.

Adding, removing, and changing the order of sample type text

- Click Preferences → Edit User Configurations or click on the upper tool bar.
   The User Configuration window appears.
- 2. Click the Vocabularies tab, then click the Sample Type tab. (Figure 268)



#### Adding a Sample Type

- 1. Click inside the Add Sample Type field, then enter your new Sample Type.
- 2. Click Add.

#### Deleting a Sample Type

- 1. Click to highlight the Sample Type you want to delete.
- 2. Click Remove.

#### Re-arranging the order currently displayed Sample Types

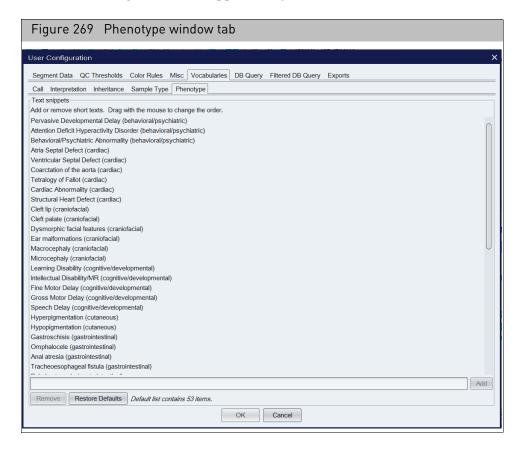
- 1. Click to highlight the Sample Type you want to move.
- 2. Drag and drop it to its new location (order).
- 3. If needed, repeat steps 1-2 to re-arrange additional Sample Types.

### Restoring the factory default Sample Types

Click Restore Defaults.
 The factory default Sample Types are restored.

Adding, removing, and changing the order of phenotype text

- Click Preferences → Edit User Configurations or click on the upper tool bar.
   The User Configuration window appears.
- 2. Click the **Vocabularies** tab, then click the **Phenotype** tab. The following window tab appears: (Figure 269)



#### Adding a short Phenotype text

- 1. Click inside the Add Phenotype Type field, then enter your new Phenotype.
- 2. Click Add.

#### Deleting a short Phenotype text

- 1. Click to highlight the Phenotype you want to delete.
- 2. Click Remove.

#### Re-arranging the currently displayed Phenotypes order

- 1. Click to highlight the Phenotype you want to move.
- 2. Drag and drop it to its new location (order).
- 3. If needed, repeat steps 1-2 to re-arrange additional Phenotypes.

#### Restoring the factory default Phenotypes

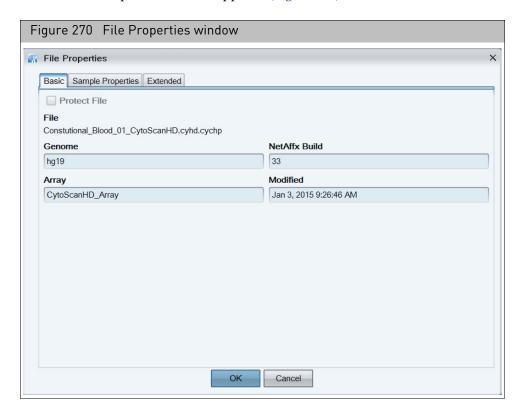
1. Click Restore Defaults.

The factory default Phenotypes are now restored.

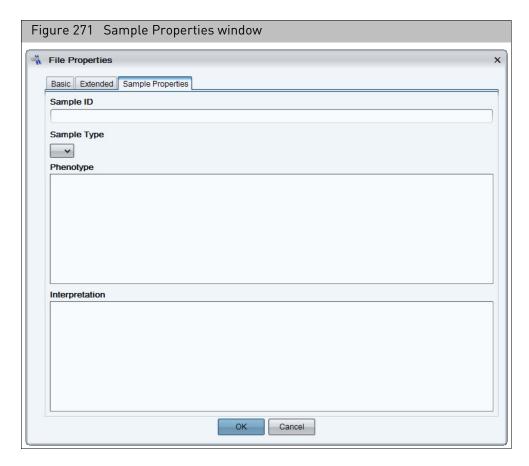
# Adding annotations at the sample (xxCHP) file level

- Right-click on a File name you want to add a Sample level annotation to.
   A menu appears.
- 2. Click View/Edit Properties.

The File Properties window appears. (Figure 270)



3. Click on the **Sample Properties** tab. (Figure 271)



- 4. Use this window to Add/Enter Sample ID(s), choose a Sample Type(s), enter Phenotype(s), and Sample Interpretation(s). **Note:** The Sample ID defaults to the File Name, but you can edit/change the Sample ID name if you want.
- 5. Click OK.

## Segment annotations

Segment level annotations such as Call, Interpretation and Inheritance can be added to segment data.

## Setting up the calls feature

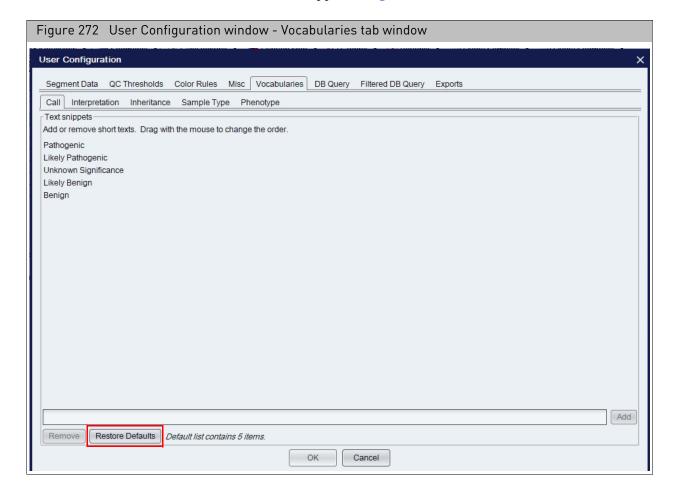
**Note:** If you are using a user profile from a previous version of ChAS, your default set of Calls will NOT appear in the Calls drop-down list. To restore them, click on Edit User Configurations  $\rightarrow$  Vocabularies  $\rightarrow$  Calls, then click the Restore Defaults button, as shown in Figure 272 on page 253.

#### Adding and removing calls

Click Preferences → Edit User Configurations or click on the upper tool bar.

The User Configuration window appears.

Click the Vocabularies tab.
 The Calls window tab appears. (Figure 272)



# Adding, deleting, and re-arranging calls

The Calls window enables you to Add, Delete and Re-arrange current Calls.

#### Adding calls to the Call drop-down list

- 1. Click inside the Add Call field, then enter your new Call.
- 2. Click Add.

Your newly added Call now appears in the Call drop-down-list.

#### Deleting calls from the Call drop-down list

- 1. Click to highlight the Call you want to delete.
- 2. Click Remove Call.

Your newly deleted Call is removed from the Call drop-down-list.

#### Re-arranging the order of Calls in the Call drop-down list

- 1. Click to highlight the Call you want to move.
- Drag and drop it to its new location (order).Your Call is now in its revised position/order within the Call drop-down-list.
- 3. If needed, repeat steps 1-2 to re-arrange additional Calls.

#### Restoring the factory default Calls

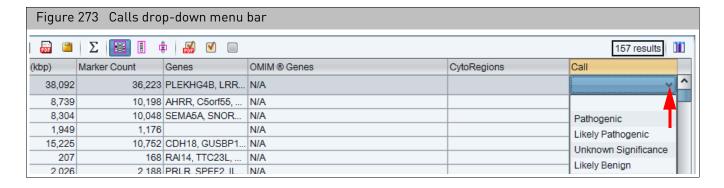
1. Click Restore Defaults.

The factory default Call(s) are now populated in the Call drop-down-list.

## Using the calls feature

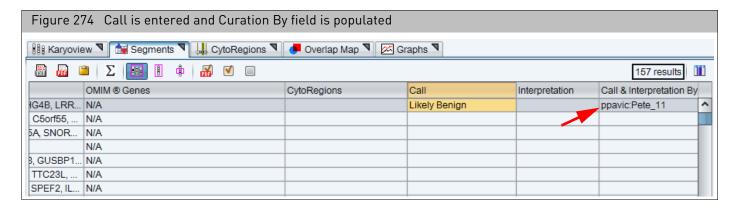
#### Method 1: At the segments table

- 1. Click the Browser's **Segments** tab.
- 2. Scroll the Segment table to the right until you see the **Call** column.
- 3. Locate the appropriate row.
- Single-click inside the field.A blue drop-down bar appears.
- 5. Click on the drop-down to reveal the list of available Calls, as shown in Figure 273



- 6. Click to select the appropriate Call.
- 7. Click outside the field or press **Enter** on the keyboard.

8. Your Call is entered. Note that the **Curation By** column is populated with the user's Windows login ID (left) and ChAS User Profile ID (right). (Figure 274)



A Call can be assigned to multiple segments at the same time. To do this:

- 1. Shift-click or Ctrl-click on the Calls fields for the segments you want to assign to the same call to.
- 2. Right-click on the highlighted area, then click **Set Value**.
- 3. Select the Call from the drop-down, then click **OK** to assign that call to all the selected segments.

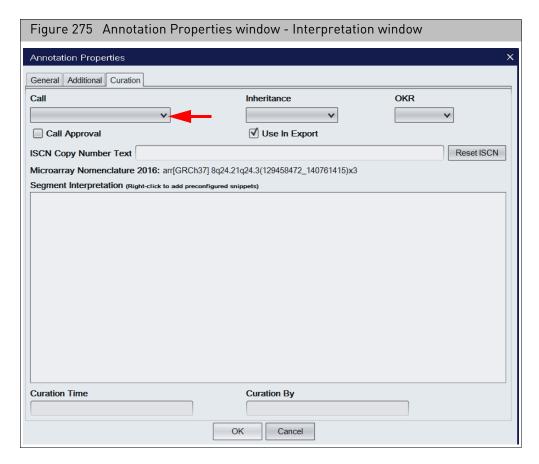
#### Method 2: At the View/Edit Annotation Properties Window

1. From any graphical view (Karyoview, Selected Chromosome View, or Details View), right-click on the segment you want to add an interpretation to, then click the menu selection **View/Edit Annotation Properties**.

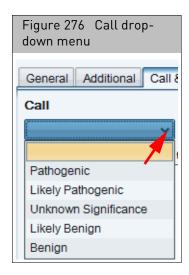
The Annotation Properties window appears.

2. Click the Curation tab.

The Interpretation window tab appears: (Figure 275)



3. Click the **Call** drop-down menu, then select your appropriate interpretation Call. (Figure 276)



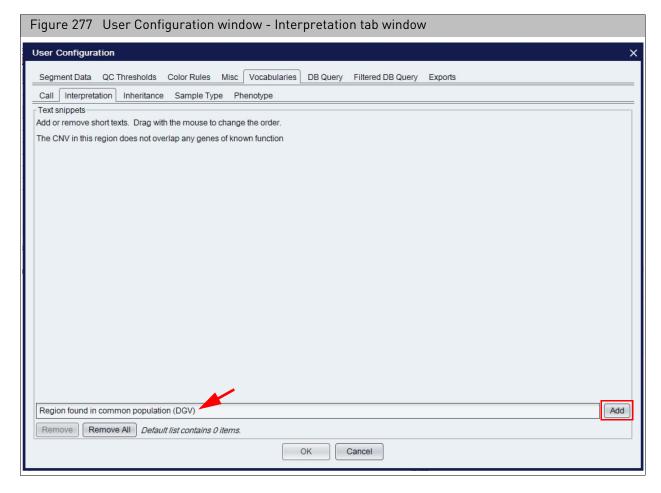
# Adding or removing interpretation snippets

The Snippets feature can be used in conjunction with the free-typing Interpretation field (below the Calls drop-down). It allows for a convenient "shortcut" when common words or phrases are used often in an interpretation.

Click Preferences → Edit User Configurations or click on the upper tool bar.

The User Configuration window appears.

2. Click the **Vocabularies** tab, then click the **Interpretation** tab. The Interpretations window tab appears. (Figure 277)



- 3. At the Interpretation window, click inside the field shown (Figure 277), then enter the snippet you want to use with your interpretation(s).
- 4. Click **Add**.

The snippet now appears and is saved in the Interpretations List pane.

- 5. Repeat steps 3 and 4 to add additional snippets.
- 6. Click OK.

#### Removing a saved snippet

- 1. Click to highlight the snippet you want to remove.
- Click Remove.

#### Removing multiple saved snippets

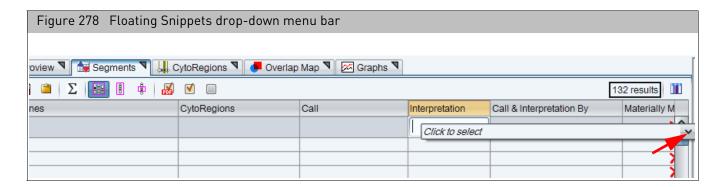
- 1. Shift click or Ctrl click to highlight each snippet you want to remove.
- 2. Click Remove All.

#### Using the interpretation snippets feature

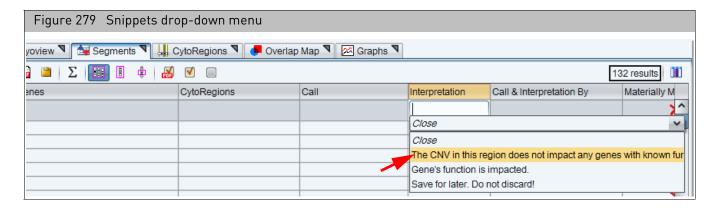
#### Method 1: At the segments table

- 1. Click the Browser's **Segments** tab.
- 2. Scroll the Segment table to the right until you see the **Interpretation** column.
- 3. Locate the appropriate row.
- 4. Click inside the field.
- 5. Right-click on the flashing cursor. If the field has existing text, place the flashing cursor to the point where you want to add a snippet, then right-click.

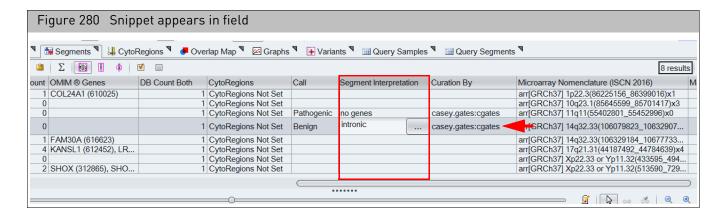
A floating drop-down menu bar appears. (Figure 278)



6. Click the snippet bar's drop-down to display the available snippets, then click on the appropriate snippet. (Figure 278)



The snippet appears in the appropriate Segment Interpretation row. (Figure 280)



7. Click outside the field.

Your snippet is entered. **Note:** The **Curation By** column is populated with the user's Windows login ID (left) and ChAS User Profile ID (right), as shown in Figure 280.

#### Method 2: At the View/Edit annotation properties window

 From any graphical view (Karyoview, Selected Chromosome View, or Details View), right-click on the segment you want to add an interpretation to, then click the menu selection View/Edit Annotation Properties.

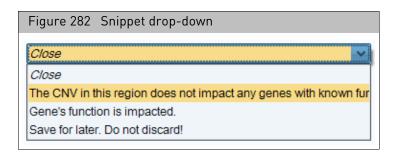
The Annotation Properties window appears.

- 2. Click the **Curation** tab.
- 3. Type in your interpretation. If at any point you want to insert your preset snippet(s), right-click inside the interpretation window.

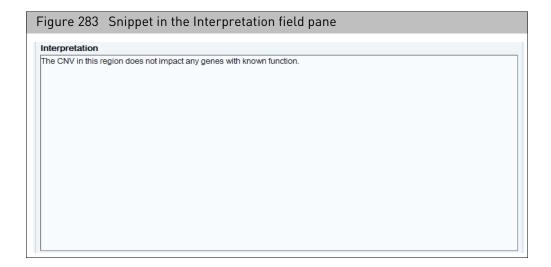
A drop-down bar graphic appears. (Figure 281)



4. Click to select the appropriate snippet or click **Close** to exit. (Figure 282)



Your snippet (preset word, sentence, or phrase) now appears "in line" with your typed text. (Figure 283)



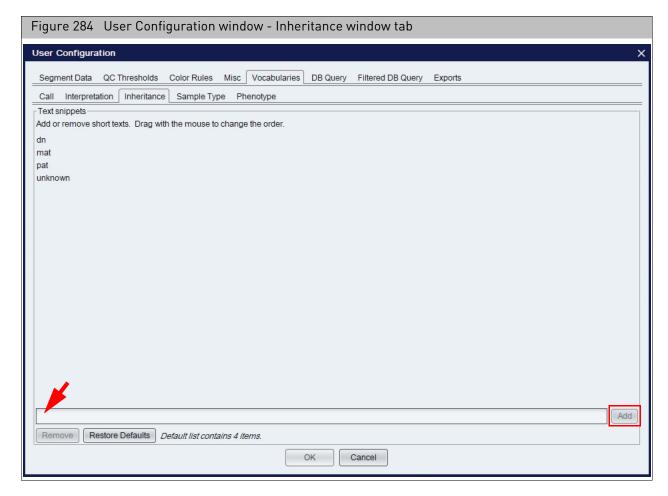
#### Adding or removing inheritance calls

1. Click **Preferences**  $\rightarrow \rightarrow$  **Edit User Configurations** or click  $\bigcirc$  on the upper tool



The User Configuration window appears.

2. Click the **Vocabularies** tab, then click then **Inheritance** tab. The Inheritance window tab appears. (Figure 284) The Inheritance window enables you to Add, Delete and Re-arrange current Inheritance calls.



 Check the Include Inheritance in Microarray Nomenclature check box to have Inheritance column entries automatically appended to the Microarray Nomenclature field.

**Example**: If the **dn** is selected in the Inheritance column then the Microarray Nomenclature field would be updated to read: **arr[GRCh37] 22q13.31(44582928\_44851090)x3 dn** 

#### Adding an Inheritance call to the Inheritance drop-down list

- 1. Click inside the Add Inheritance Call text field (Figure 284), then enter your new call.
- 2. Click Add.

#### Deleting a call from the Inheritance drop-down list

- 1. Click to highlight the call you want to delete.
- 2. Click Remove.

#### Re-arranging the order of calls displayed in the Inheritance drop-down list

- 1. Click to highlight the Inheritance you want to move.
- 2. Drag and drop it to its new location (order).

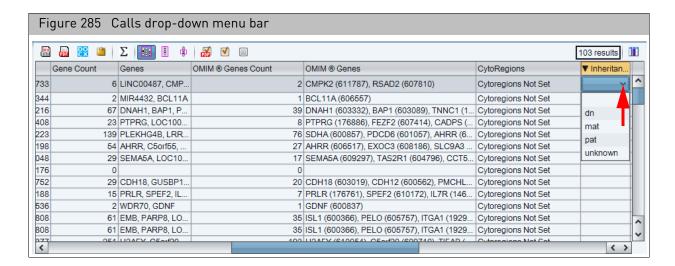
Your call is now in its revised position/order within the Inheritance drop-downlist.

3. If needed, repeat steps 1-2 to re-arrange additional Inheritance.

## Using the inheritance feature

#### Method 1

- 1. Click the Browser's **Segments** tab.
- 2. Scroll the Segment table to the right until you see the Inheritance column.
- Single-click inside the field.A blue drop-down bar appears.
- 4. Click on the drop-down to reveal the list of available Inheritance Calls. (Figure 285)



- 5. Click to select the appropriate Inheritance.
- 6. Click outside the field or press **Enter** on the keyboard. Your Inheritance is entered.

#### Method 2: From the View/Edit annotation properties window

 From any graphical view (Karyoview, Selected Chromosome View, or Details View), right-click on the segment you want to add an interpretation to, then click the menu selection View/Edit Annotation Properties.

The Annotation Properties window appears.

- 2. Click the Curation tab.
- 3. Select the Inheritance call from the drop down.

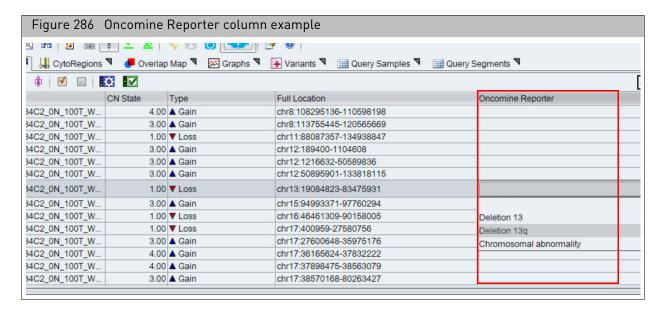
### Adding Oncomine Reporter annotations

You may add annotations based on approved Oncomine Reporter search terms. Once your segments have been annotated with Oncomine Reporter term(s), the Segments Table can be exported as a txt file and directly uploaded to Oncomine Reporter for literature searches based on the assigned annotation.

The annotation options for use with Oncomine Reporter are a controlled vocabulary.

**Note:** The options available in the drop-down list for the Oncomine Reporter column are compatible nomenclature with Oncomine Reporter application.

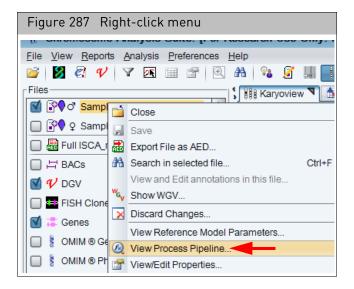
- 1. Select to view the Oncomine Reporter column in the Segments Table. See "Selecting columns to display or hide" on page 329.
- 2. Click in the Oncomine Reporter column for a given segment to assign the appropriate annotation for that segment. Segments on a given chromosome will only see relevant Oncomine Reporter terms for that chromosome. For example, if there is a gain of Chromosome 8q, you will only see Oncomine Reporter terms for Chromosome 8. All other terms for other chromosomes are hidden for that segment.



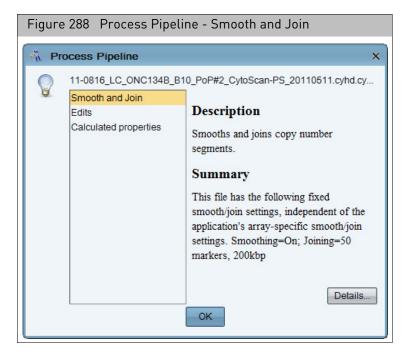
For information on Oncomine Reporter Table State, "Saved table states" on page 332 on Table States.

## Tracking and reviewing the log file

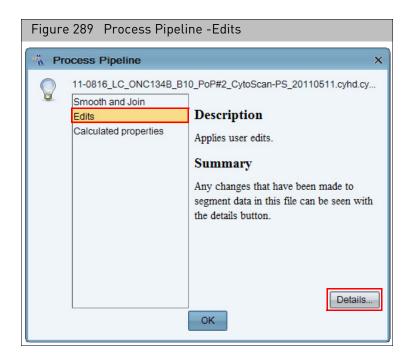
1. Right-click on the file, then click **View Process Pipeline**. (Figure 287)



The **Process Pipeline** window appears. (Figure 288)

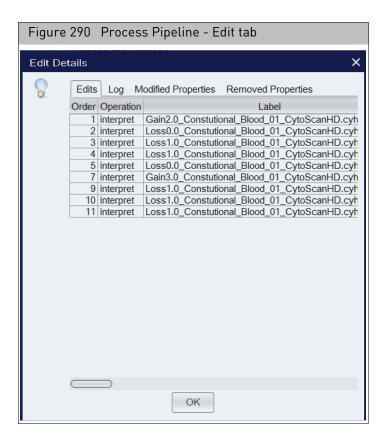


- 2. Click **Smooth and Join** for its description and summary. (Figure 288)
- 3. Click **Edits** for its description and summary of the file's past user edits. (Figure 289)



#### 4. Click Details.

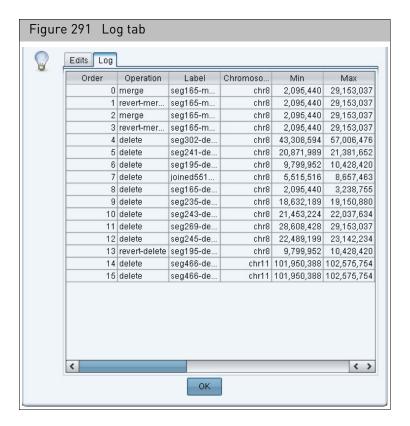
A Edit Details window appears featuring an Edit and Log tab. (Figure 290)



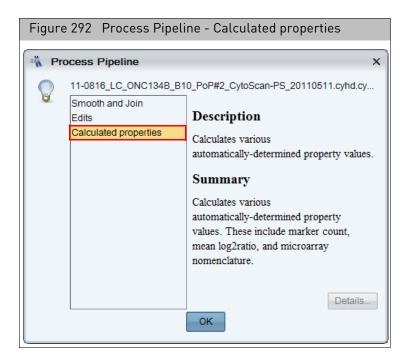
The **Edit** tab lists only recent edits that were NOT reverted. Use the horizontal scroll bar to reveal additional edit information. For a complete historic account of all edits made (including reversions of some edits), click on the **Log** tab.

5. Click **OK** to return to the **Process Pipeline** window or click the **Log** tab.

The **Log** tab displays an extensive summary of the file's edited history to date. Use the horizontal scroll bar to reveal additional log information. (Figure 291)



6. Click **OK** to return to the Process Pipeline window. (Figure 292)



- 7. Click **Calculated Properties** for its description and summary.
- 8. Click **OK** to return to the ChAS browser.



## Using CytoRegions

#### This chapter includes:

- "CytoRegions overview"
- "Selecting a CytoRegions information file" on page 269
- "Viewing CytoRegions" on page 271
- "Searching CytoRegions" on page 275
- "Using filters with CytoRegions" on page 276
- "Using restricted mode" on page 276
- "Assigning a CytoRegion for targeted XON analysis" on page 278
- "Creating an AED File from a gene list" on page 278

## CytoRegions overview

The CytoRegions feature enables you to define parts of the genome that are of special interest to you.

**Note:** The CytoRegions feature is designed 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 Genes, is not recommended for use as a CytoRegions file due to the large number of reference annotations.

To use CytoRegions, you need to select a file(s) with position information for regions of the genome as the CytoRegions file.

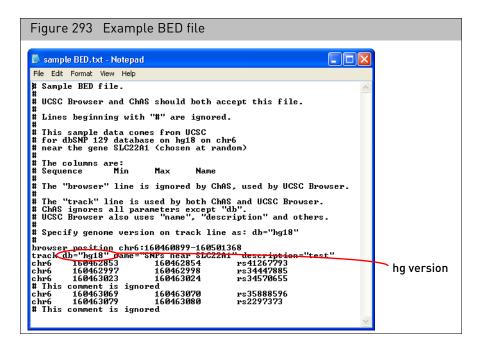
- Select Region information files in AED or BED format.
- Use existing Region file(s), or create a new one in AED format in ChAS, then add regions to it by selecting segments, annotations, or regions in other loaded files.
- Add annotations to regions to help you track the information. See "Creating and editing AED files" on page 288.

After selecting a CytoRegions file, you can:

- Use the Restricted Mode to display only Segments and graph data that appear in those regions. While in this mode, 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.
- Protect a CytoRegions file. See "Protecting an AED file" on page 309.

## Selecting a CytoRegions information file

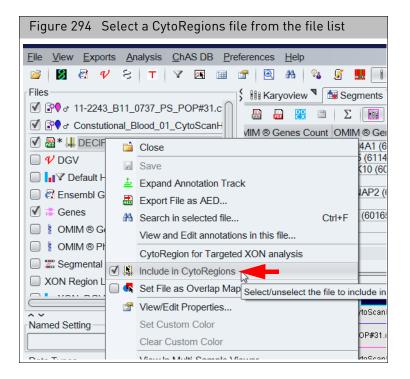
Select the CytoRegions file from the available region information files. See "Loading files" on page 122. The software automatically checks the hg version of an AED or BED file before loading. See Figure 293 for an example BED file. The file will not be loaded if the hg version does not match what is loaded in the ChAS Browser. If an hg version is not found for the AED or BED file, a warning message appears.



Do either of the following to select a CytoRegions file:

- In the files list, right-click a file and select **Include as CytoRegion** on the shortcut menu, as shown in Figure 294.
- Multiple files can be selected to be included in the CytoRegions. To do this, rightclick on each AED/BED file to be included, then select Include in CytoRegions.

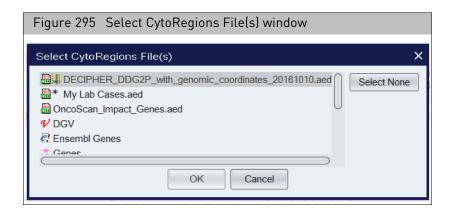
**Note:** All files (included in CytoRegions) are treated as they are one file.



#### Or

From the menu bar, click View → Select as CytoRegions file(s).
 Alternatively, in the CytoRegions Tab of the Upper display, click the Select to include in CytoRegions button.

The Select CytoRegions File(s) window opens. (Figure 295)



Select a regions information file(s) to use for CytoRegions, then click OK.

**Note:** You may select more than one Region file (AED/BED) to include as CytoRegions. ChAS handles multiple selected files as single (larger) CytoRegion file.

The **Create New** feature is described in "Creating an AED file of annotations" on page 288.

**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 after the software is opened the next time with the same user profile. To clear a CytoRegions file map from ChAS, click **Select None** (Figure 295). Alternatively, right-click the file in the Files list, then deselect **Include in CytoRegions** on the shortcut menu to toggle the check mark off.

## Viewing CytoRegions

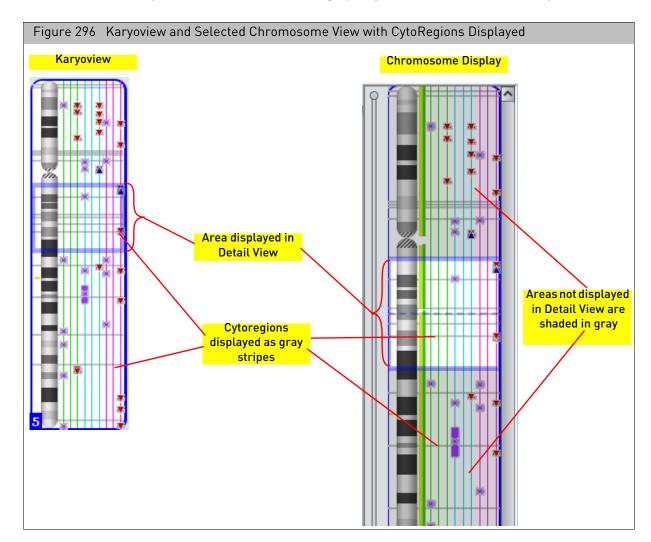
After selecting a CytoRegions file, it can be displayed in either a graphic view or table.

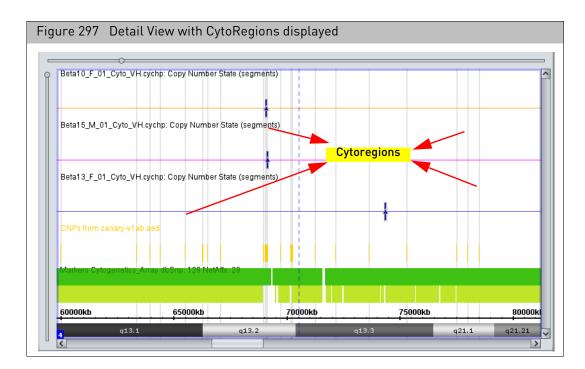
- "CytoRegions in the graphic views" on page 271
- "CytoRegions table" on page 273

**Note:** CytoRegions that share genomic coordinates with a detected segment are listed in the "CytoRegions" column of the Segments table. See "Segments table" on page 338.

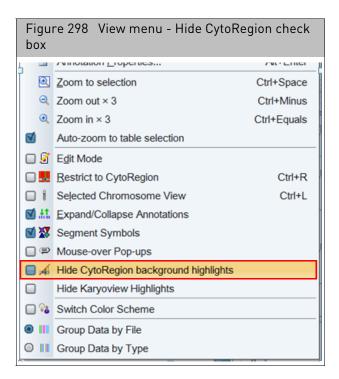
## CytoRegions in the graphic views

Regions specified in the CytoRegions file are displayed as dark gray stripes in the Karyoview and Chromosome Display (Figure 296) and Detail View (Figure 297).





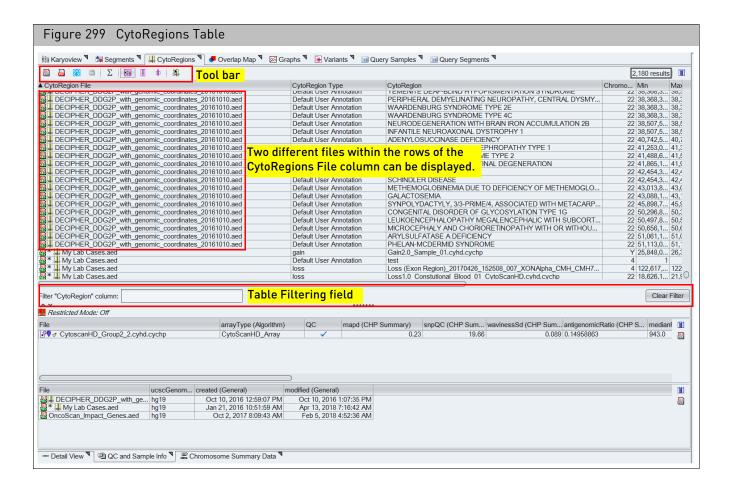
You can turn off highlights by un-checking the View menu's **Hide CytoRegion Background Highlights** check box. (Figure 298) This action keeps the CytoRegions in the Table View, however they are no longer highlighted in the Graph Views.



#### CytoRegions table

The CytoRegions table (Figure 299) 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 copynumber 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.



#### 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.

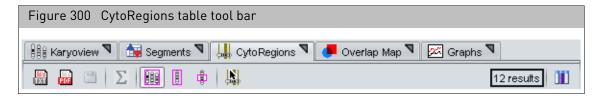
 Click a row in the CytoRegions table to select the corresponding annotation from the CytoRegions file. All of the lines for that region are highlighted in the table.
 The Details View zooms to the currently selected region.

**Note:** If the Details View does not automatically zoom to the selected region, confirm that the Auto-zoom option is selected (click  $View \rightarrow Auto-zoom$  to table selection from 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 tool bar button in the CytoRegions table to show only the cytoregions in the current region.

#### CytoRegions tool bar

The tool bar (Figure 300) provides quick access to table functions. The standard functions are described in "Standard tool bar controls" on page 328.



The tool bar has one specialized button.



Select CytoRegions file (See "Selecting a CytoRegions information file" on page 269 )

CytoRegion Table Column	Description
CytoRegion File	Name of the AED/BED file that contains the region.
CytoRegion Type	Type of file or element from which the CytoRegion is derived. Default User Annotations are annotations derived from AED or BED file.
CytoRegion	Identifier for region.
Chromosome	Chromosome in which the region is located.
Min	Starting position of the region.
Max	Ending position of the region.
Size (kbp)	Size of the region.
Segment Label	Label assigned to the detected segment by ChAS.
Segment Name/ID	Name/ID of the Segment.
Segment File	Sample File that the segment was detected in.

CytoRegion Table Column	Description
Segment Type	Type of segment:  - CN loss or gain  - XON loss or gain  - Mosaic loss or gain  - LOH
Segment Min	Starting position of segment.
Segment Max	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 Appendix E, "Genomic position coordinates" on page 487).
Segment Size (kbp)	Size of the segment.
CN State	Copy Number State (Displayed for Gain, Loss, and Mosaicism segment types).
Shared Size	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" on page 327.
- Export data in various formats. See "Exporting table data" on page 420..
- Search CytoRegions to display only regions of interest (see below).

## **Searching CytoRegions**

The Search CytoRegions feature enables you to search the "CytoRegion" column for text strings that match a search string.

#### 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 enables 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.

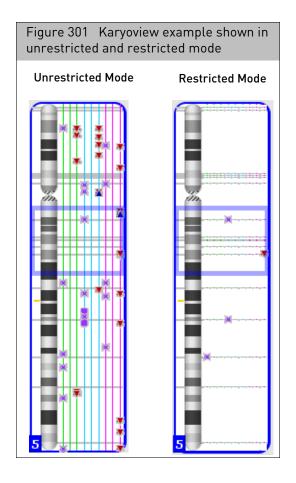
To select/deselect Restricted Mode:

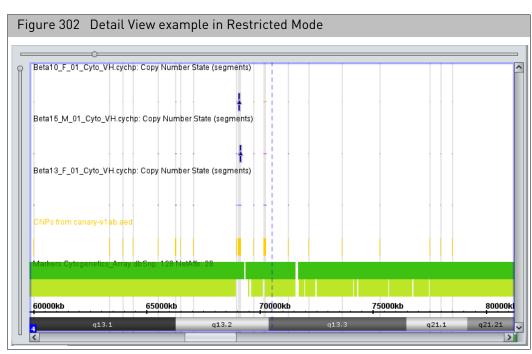
From the View menu, select Restrict to CytoRegion.

Or

• Click the **Restricted Mode** button from the main tool bar.

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 301, Figure 302)





When you deselect Restricted Mode, an Exit Restricted Mode window appears. (Figure 303)



Click **Yes** to exit Restricted Mode.

## Assigning a CytoRegion for targeted XON analysis

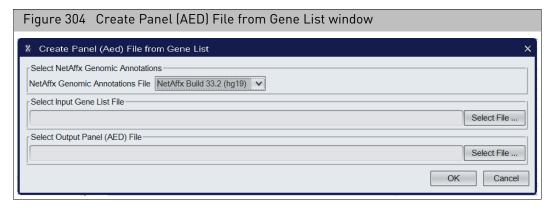
This option is designed for use with CytoScan XON array data. Assigning a CytoRegion for a Targeted XON analysis combines these two steps:

- 1. Assigns the file as the CytoRegion.
- 2. Automatically sets the appropriate filters for Targeted XON analysis.
  - Hides all segments in the Genome region and colors all the data gray.
  - Turns on all levels in CytoRegions to reveal any XON Region segment call and colors the data the assigned sample color.

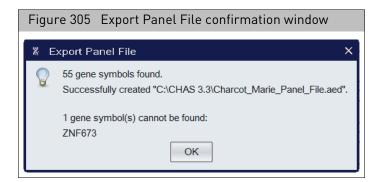
## Creating an AED File from a gene list

To create an AED file from a Gene List for use with the CytoRegion for Targeted XON analysis or other region file (CytoRegion or Overlap Map) do the following:

- Create a text tab delimited file containing your list of Genes for which you would
  like to generate an AED file. The gene coordinates will be pulled from a selected
  NetAffx Genomic Annotation file. The Gene List should just be a list of gene
  names with no header and saved as a text tab delimited file.
- Click Exports → Create and Export an AED file from a Gene List.
   The following window appears:



- 3. Select the NetAffx Genomic Annotation file from the drop down list. The annotation database files visible are available files from the Library file folder. If you do not see a NetAffx Genomic Annotation file that you want to use, please copy the file into C:\Affymetrix\ChAS\Library and restart the ChAS browser.
- 4. Select an Input Gene List File: Click on Select File button and browse to the location of your TXT tab-delimited Gene List file.
- 5. Select an Output AED File: Click on the Select File button, select the location to save the output file, name the AED file, click **Save**.
- 6. Click **OK** to generate the AED file. (Figure 305)



**Note:** If entries in the Gene List are not found in the NetAffx Genomic Annotation file, they will be listed in a message box. Any skipped entries can be added to the AED file. See "Creating and editing AED files" on page 288 and "Adding annotations to an AED file" on page 290.



## Using the overlap map and filter

#### This chapter includes:

- "Overlap map and filter overview"
- "Using the overlap filter"
- "Selecting the overlap map file"
- "Viewing overlap regions" on page 283
- "Using the overlap filter" on page 287

## Overlap map and filter overview

- The Overlap Map enables you to show or hide segments in areas of the genome that you are not interested in. For example, in regions of Benign Copy Number Polymorphism, you can:
  - Specify regions of the genome in an Overlap Map File.
  - Optionally filter out detected segments that are overlapped by these regions.
     You can also 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.
- Protect an Overlap Map file. See "Protecting an AED file" on page 309...

## Using the overlap filter

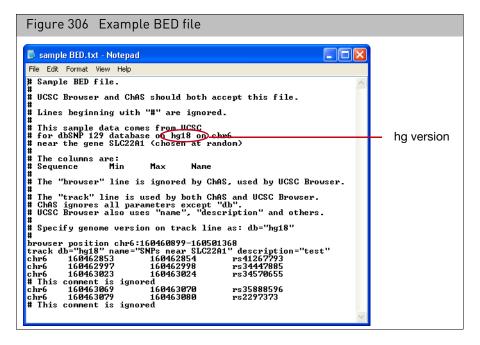
- 1. Select a region information file for the Overlap Map. For supported file types, see "Selecting the overlap map file" (below).
- 2. Set Percent Overlap value for each segment type you want to filter in the Segment Filters window. See "Using the overlap filter" on page 287.
- 3. Choose how you want to view the Overlap Map regions.
  - "Viewing overlap map regions in the graphic displays" on page 283
  - "Viewing the overlap map table" on page 284

## Selecting the overlap map file

Use the following file types for the Overlap Map:

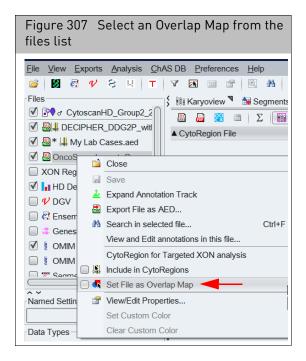
- Region information files in AED and BED format.
- Position information from Reference annotation files.

The software automatically checks the hg version of an AED or BED file before loading (see Figure 306 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 appears.



Do either of the following to select an Overlap Map:

1. In the files list, right-click the file and select **Set File as Overlap Map** on the shortcut menu. (Figure 307)

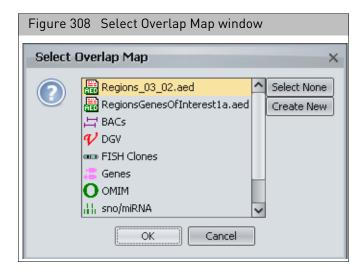


#### OR

1. On the menu bar, click Select View  $\rightarrow$  Overlap Map.

Alternatively, in the Overlap Map tab in the upper display, click the Select Overlap Map tool bar button.

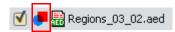
The Select Overlap Map window appears. (Figure 308)



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

Select the file you want to use, then click OK.
 The region file loads and its regions are displayed with overlap information in the Overlap Map tab.

The Overlap Map icon \_\_ appears next to your selected file within the File List, as shown below.



**Note:** To clear an Overlap map, click **Select None**. (Figure 308) Alternatively, right-click the file in the files list, then select **Set File as Overlap Map** from the shortcut menu.

## Viewing overlap regions

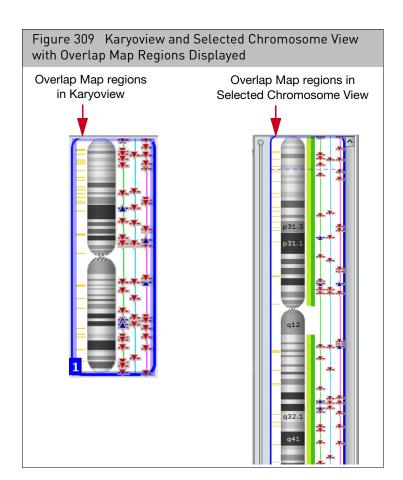
You can view the Overlap Map regions:

- In the graphic display views. See "Viewing overlap map regions in the graphic displays" on page 283.
- In the Overlap Map window tab. See "Viewing the overlap map table" on page 284.

**Note:** Overlap Map items that are covered by a detected segment are listed in a column of the Segments table. "Segments table" on page 338.

Viewing overlap map regions in the graphic displays

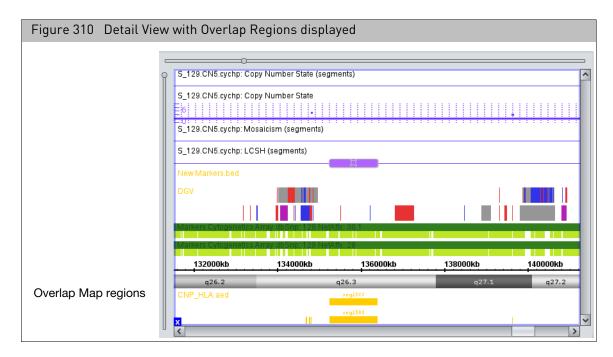
In the Karyoview and Selected Chromosome View, regions specified in the Overlap Map file are displayed as short rectangles to the immediate left of the Cytobands. (Figure 309)



In the Detail View, the overlap map regions are displayed as rectangles below the Cytobands (Figure 310). 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" on page 300.
- Using the Color Rules. See "AED/BED color rules" on page 310.



## Viewing the overlap map table

The Overlap Map table (Figure 311 on page 285) 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.

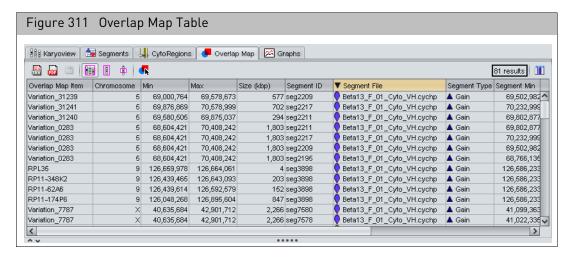
#### Highlighting overlap regions in the overlap map table and details view

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 (click **View** → **Auto-zoom to table selection** from 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 tool bar button in the Overlap Map table to show only the overlaps in the current region.



#### Overlap map tool bar

The tool bar provides quick access to table functions. The standard functions are described in "Standard tool bar controls" on page 328.





Select Overlap Map file (see "Selecting the overlap map file" on page 281).

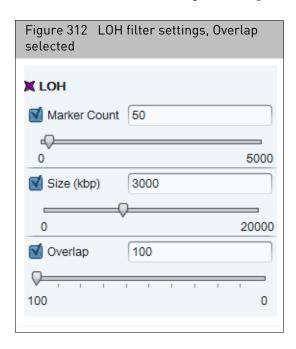
Overlap Map Table Column	Description
Map Item Type	Source of the position information (CN Gain or Loss segment, reference annotation, etc.)
Overlap Map Item	Identifier for the item.
Chromosome	Chromosome in which the item is located.
Min	Starting position of the item.
Max	Ending position of the item.
Size (kbp)	Size of the item.
Segment Label	Label assigned to the detected segment by ChAS.

Overlap Map Table Column	Description
Segment Name/ID	Name/ID of the Segment.
Segment File	Sample File that the segment was detected in.
Segment Type	Type of segment: CN loss or gain Mosaicism LOH
Segment Min	Starting position of segment.
Segment Max	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 Appendix E, "Genomic position coordinates" on page 487).
Segment Size	Size of the segment.
CN State	Copy Number State (Not displayed for other segment types).
% Overlap	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.
% Coverage	How much of the Overlap Map item is covered by the Segment.
Shared Size	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.

## Using the overlap filter

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

- 1. Select a region information or Reference Annotation file for the Overlap Map.
- 2. Click the **Y** button to open the **Segment Filters** window. (Figure 312)



- 3. Select the Overlap check box(es) for the segment types you want to filter.
- 4. Use the slider to set the parameter's value or enter a value in the adjacent text box.
  - As you move the slider farther to the right (or enter smaller values in the box)
     more and more of the Overlapped segments are 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 a setting is 1%.
  - The results of filtering are seen instantly in all tables and graphs.
  - See "Using filters with CytoRegions" on page 276. for information about the interactions of the Overlap Map filter with the CytoRegions.



## Creating and editing AED files

#### This chapter includes:

- "Creating an AED file of annotations" on page 288
- "Creating an AED file of regions" on page 291
- "Viewing or Editing AED File Properties" on page 295
- "Viewing and editing annotations" on page 298
- "Viewing and batch editing AED file contents" on page 307
- "Protecting an AED file" on page 309
- "AED/BED color rules" on page 310
- "Exporting information in AED or BED format" on page 318

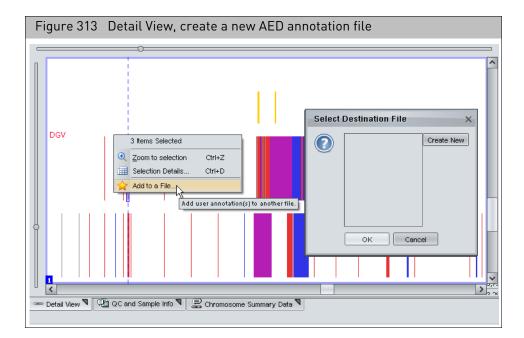
## Creating an AED file of annotations

You can create AED files that contain:

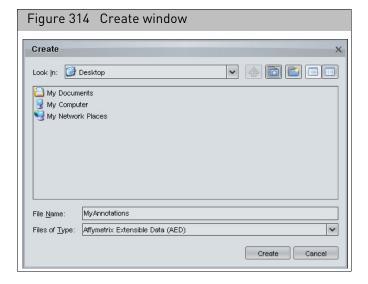
- User-selected annotations. The annotations in an AED annotation file can be edited in ChAS.
- Position information on regions of the chromosome, as well as additional annotation information on the regions. **Note:** AED region information files can be used for CytoRegions or Overlap Map functions.
- Third party reference annotations converted to AED file format.
- User-generated AED region information files can be used for CytoRegions or Overlap Map functions. Note: You can add a region to a new or previously created AED file by selecting these feature types:
  - Detected Segments
  - Reference Annotations
  - Regions in previously loaded files
- 1. In the Detail View, select the non-AED annotation(s) for the AED file using one of the following methods:
  - Right-click an annotation and select **Add to a File** on the shortcut menu. (Figure 313)

Or

- Draw a box around multiple annotations, right-click the selection, and select **Add to a File** on the shortcut menu.
- 2. In the window that appears (Figure 313), click Create New.



3. In the Create window (Figure 314), click to select a folder, then enter a file name.



#### 4. Click Create.

The **Select Destination File** window appears and displays the name of your new AED file. (Figure 315)



#### 5. 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

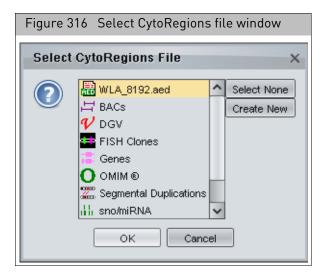
- 1. In the Detail View, use one of the following methods to select the annotation(s) you want to add to an AED file:
  - Right-click an annotation and select Add to a File on the shortcut menu OR
  - Draw a box around multiple annotations, right-click the selection, and select **Add to a File** on the shortcut menu.
- 2. In the **Select Destination File** window, select an AED file, then 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

**Note:** You may want to incorporate new segments of features from a set of samples in a new Regions file. 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" on page 318.

Creating 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 window appears. (Figure 316)

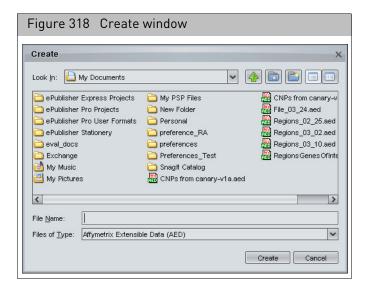


**Note:** You can also create a new AED file when adding a region to an AED file.



2. Click **Create New** in the Select File window.

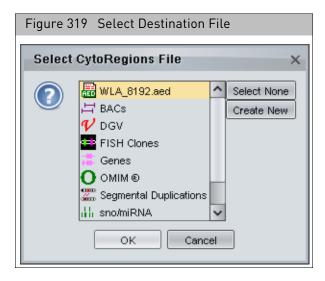
#### The Create window appears. (Figure 318)



- 3. Use the window controls to browse to a folder for the AED file.
- 4. Enter a file name.
- 5. Click **Create** in the Create window.

The Select File window appears with the new file selected. (Figure 319)

**Note:** The AED file is automatically assigned the same genome assembly version (i.e. "hg18", "hg19", etc.) as the currently loaded NetAffx annotations.



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  $\nearrow$  button or select **File**  $\rightarrow$  **Open** on the menu bar.

#### Adding regions to an existing AED file

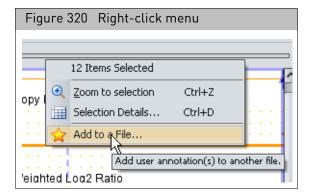
You can add a new region to an existing region information file in AED format by selecting the following features in the ChAS graphic views:

- 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 "Creating a color rule" on page 312). Alternatively, you can edit the color of a particular DGV annotation added to an AED file, in the Detail View (see "Entering general information" on page 300).

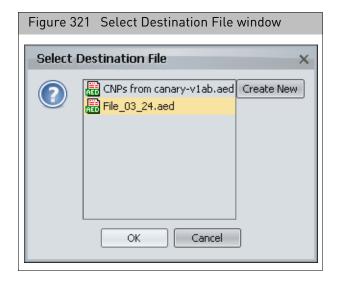
#### Adding a section to a new region (AED) file

- 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 (Figure 320):
  - 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 (see E"Expanding and contracting annotations"
    on page 197).
  - Region



The Select Destination File window appears. (Figure 321)

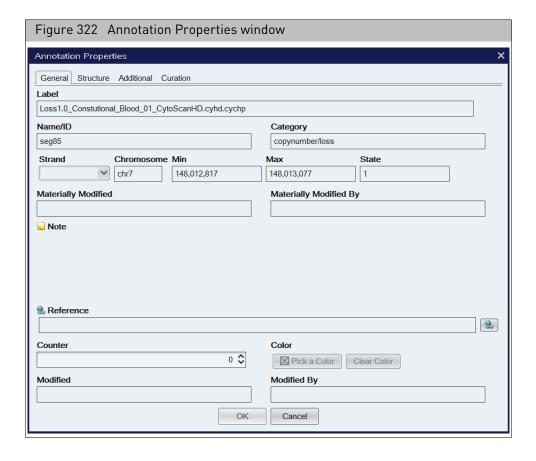
**Note:** Some options may not be available, depending upon the number of type of items you have selected.



The Select Destination File window displays a list of the currently existing AED files to which you may add the segment.

2. Select the region file you want to use, then click **OK**.

**Note:** The **Annotation Properties** window opens if you have selected a single item (Figure 322). If you have selected multiple items, the **Annotation Properties** window does not open.

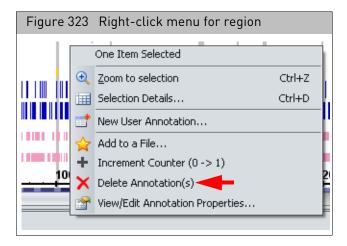


See "Viewing and editing annotations" on page 298 for information on editing the annotations.

After editing the annotation properties, click **OK**. The section is saved in the Region file.

## Deleting regions from an AED file

1. Right-click on a region in a region file and select **Delete Annotations(s)** (Figure 323).



The region is deleted from the Region (AED) file.

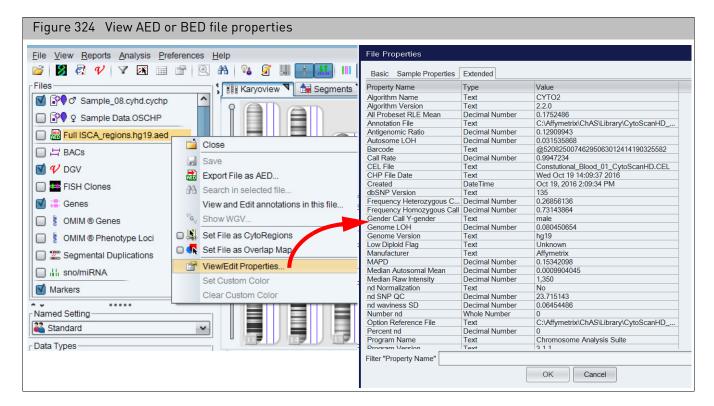
### Viewing or Editing AED File Properties

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 "Editing BED files" on page 489.

## Viewing file properties

- Right-click on a file, then click to select View/Edit Properties.
   The File Properties window appears.
- Click on the Extended tab.A table with the file properties appears. (Figure 324 on page 296)



## Viewing the genome assembly version

The assigned genome assembly version of a loaded file can be viewed in the Properties box (Figure 324). 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.

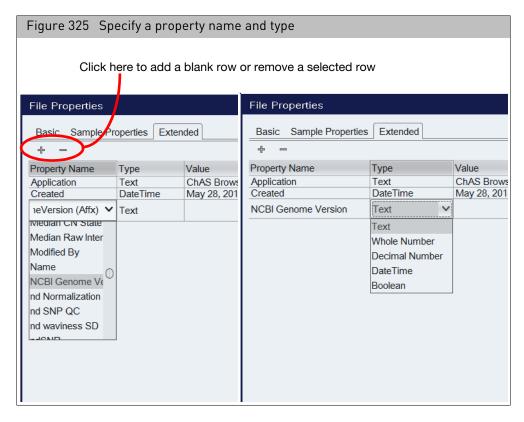
If an AED file does not include a genome assembly version, you can manually set it. To do this, in the Properties window:

- 1. Click + to add a new property row, as shown in Figure 325 on page 297.
- 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, "Constitutional").

**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 "Editing a property value" on page 298.

#### Adding a property

- 1. In the **Properties** window, click the + symbol to create a blank row (Figure 325).
- 2. In the **Property Name** field, enter a property name or make a selection from the Property Name drop-down list. (Figure 325)



- 3. Click **Type**, then make a selection from the drop-down list.
- 4. Click Value

A cursor appears inside the Value field. (Figure 326)



5. Enter a value, then click **OK**. For significantly longer values, click (Figure 326) to open a Value editing window. Enter your (longer) value in this window, then click **OK**.

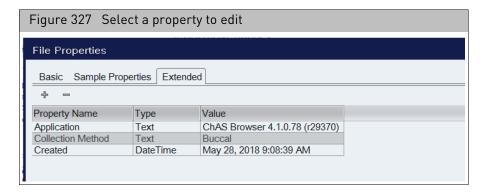
The new value is entered in the File Properties table.

6. Click OK.

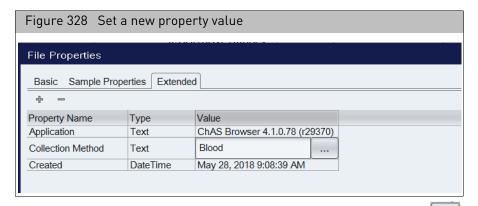
# Removing a property

# Editing a property value

- 1. In the Properties window, select the row that you want to delete.
- 2. Click the = icon.
- 1. Right-click the file and select **View/Edit Properties** on the shortcut menu. The File Properties window appears. (Figure 327)
- 2. Click on a row to select it.



For example, click the **Value** field. (Figure 328).

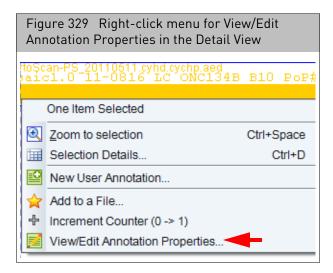


- 3. Enter a value, then click **OK**. For significantly longer values, click \_\_\_\_\_ to open a Value editing window. Enter your (longer) value in this window, then click **OK**. The new value is entered in the **File Properties** table.
- Click OK.

### Viewing and editing annotations

The Annotation Properties window opens:

- When adding a region to a Region (AED) file
- When you select View/Edit Annotation Properties from the right-click menu (Figure 329) for the following types of features:
  - Detected Segments
  - Reference Annotations (view annotations only)
  - AED Annotations (view and edit annotations)



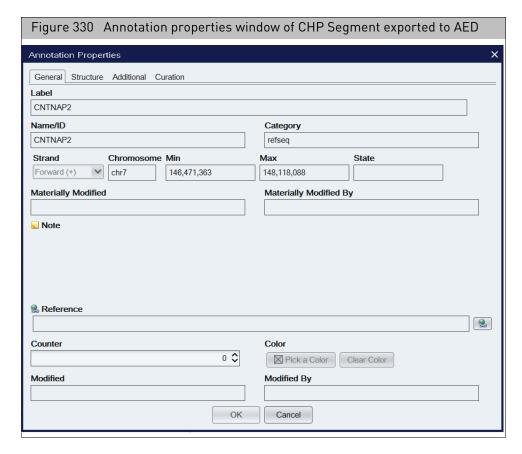
**Note:** The **View/Edit Annotation Properties** menu option is not available if you have more than one feature selected in the Detail View.

The Annotation Properties window has three tabs:

- **General** (See "Entering general information" on page 300)
- Additional (See "Adding Properties" on page 301)
- Curation (See "Adding a curation (Optional)" on page 304)

You can also create new user annotations if you select an element. This feature enables you to create a region that is not based on a segment or reference annotation. See "New user annotations" on page 305.

## Entering general information



### Editing the General tab properties in AED annotations

1. In the General window tab (Figure 330), enter, edit, or select:

Annotation	Description
Label	Label given to the AED, CHP, or other annotation. For annotations originating from CHP segments, can be the Type, State, and Filename of the CHP file. The Label is not editable.
Name/ID	Name/ID assigned to the annotation in the AED, CHP or other file.
Category	Information on the source of the region. If the region was added by selecting a CHP segment, the segment type is saved.
Strand	The Sequence Strand of the item.
Chromosome	Cannot be edited in Annotation Properties box. See "New user annotations" on page 305.
Min	The smallest of the annotation's chromosomal coordinates.
Max	The largest of the annotation's chromosomal coordinates.
Materially Modified Time	Time stamp of when the start or end coordinate, type, or state of the segment or BED/AED annotation was last altered.
Materially Modified By	The Operating System user and ChAS User IDs of the Modifier who last changed a Material property (start or end coordinate, type, or state) of the segment or annotation.
Note	Information and comments about the region.  Note: Always use alphanumeric characters and underscores. Avoid the use of odd characters. (Examples: & + ( ) [ ] { } ~ ^ and commas.)

Annotation	Description
Reference	URL/web address associated with the annotation.
	Click to link directly to the Reference from the Annotation Properties window. Internet connection is required.
Modified	The time stamp of the last modification of the annotation.
Modified By	The Operating System user and ChAS User IDs of the Modifier who last changed the annotation.
Counter	enables you to track the number of times something has been seen.
Color	Allows assignment of a hard-coded color to the Annotation in ChAS's graphical views.

# Customizing properties

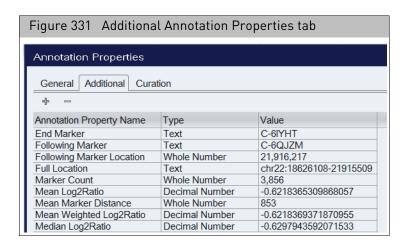
#### **Adding Properties**

In the **Additional** tab you can enter new annotation information for an AED annotation. The information will be displayed in the:

- Tool tip when you position the mouse arrow vover an annotation in the Details View
- · Selection Details window

#### Adding custom properties

1. In the Annotation Properties window, click the **Additional** tab. (Figure 331 on page 301)



Additional Annotation	Description
Annotation Property Name	Name assigned to the property.
Туре	Type assigned to the property.
Value	Value assigned to the property.

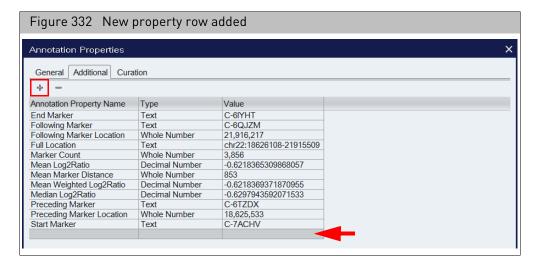
**Note:** Different default properties may already be assigned to the annotations added from reference annotations.

2. Click the button at the top of the table.

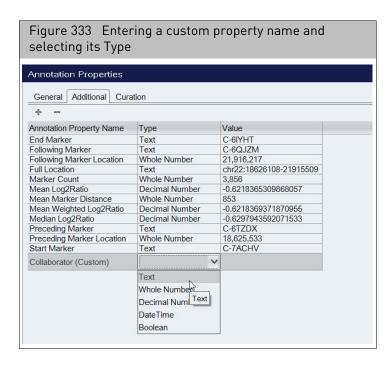
A new row appears in the table. (Figure 332)

If needed, click on a column's edge, then drag it (right) to make it wider.

You can delete a property by selecting a row, then clicking the button



3. Click in the row under the Property Name column and enter a name for the property. Note that your new entry is followed by the text "(custom)". (Figure 333) For more details, see "ChAS properties and types" on page 481.

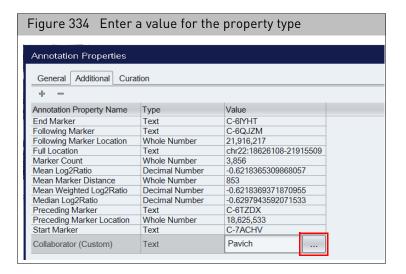


4. Click in the row under Type and select a property type from the drop-down list (Figure 334). For more details, see "ChAS properties and types" on page 481.

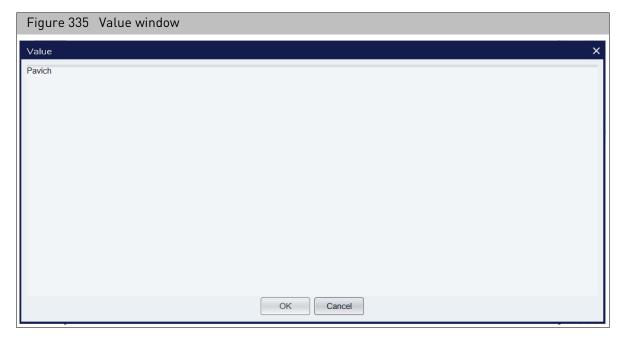
#### Choose from:

- Text
- Whole Number

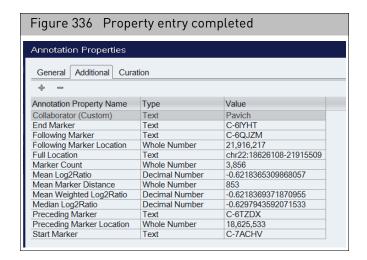
- Decimal Number
- Date Time
- Boolean
- 5. Click in the row under Value and enter the value (Figure 334).



Alternatively, click the **Browse** button (Figure 334), then in the **Value** window (Figure 335) that appears, enter the property value, then click **OK**.



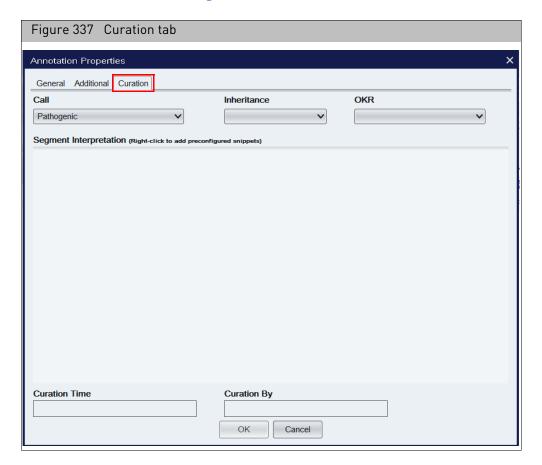
The property entry is completed. (Figure 336)



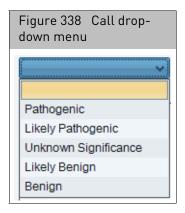
6. Click **OK** in the Annotation Properties window.

#### Adding a curation (Optional)

1. Click the **Curation** tab. (Figure 337)



2. Click the Call drop-down to select a Call. (Figure 338)

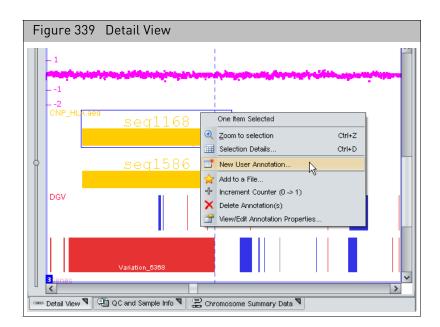


- 3. Click to select an appropriate Call.
- 4. If needed, (in the Interpretation field) type in your interpretation or right-click to add a pre-configured snippet.
- 5. Click the **Inheritance** drop-down menu to select an Inheritance mode.
- 6. Click on the OKR drop-down menu to select an OKR annotation.
- 7. Click OK.

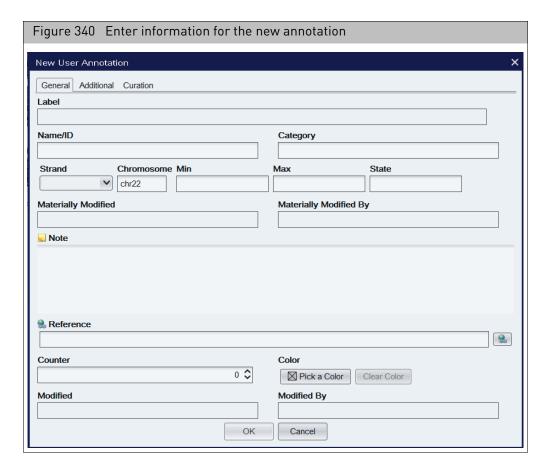
### New user annotations

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 339)



2. In the window that appears, enter the annotation information (Figure 340). For more details, see "Entering general information" on page 300 and "Customizing properties" on page 301.



#### 3. Click OK.

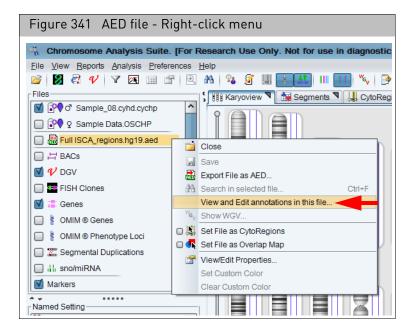
The new user annotation is created and saved in the AED file.

**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.

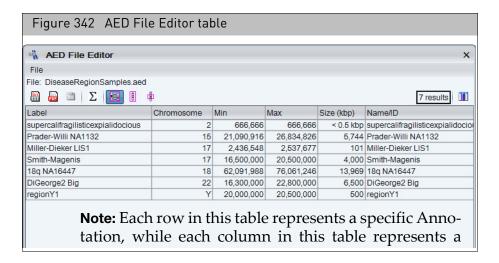
### Viewing and batch editing AED file contents

The AED Editor enables you to view multiple AED annotations and their properties in a table format and edit the values of properties easily - for single annotations or to edit a property for multiple annotations in a batch - to a new value.

At the Files pane (top left), right-click on the AED file you want to view/edit.
 A pop-up menu appears. (Figure 341)



Click View and Edit annotations in this file.
 The AED File Editor table window appears. (Figure 342)

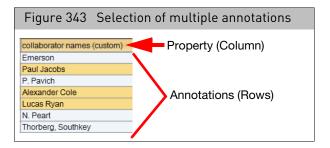


**Note:** The AED File Editor table displays ALL properties and tabs of every annotation contained in an AED file.

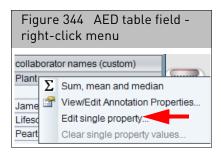
Use the scroll bar to move the table to the right to see more column entries and/or click, then drag the AED File Editor table window to make it wider.

To locate your table data faster, click on any appropriate column header to sort in a descending or ascending order. Also, the number of columns displayed in the Select Columns window, varies with each AED file.

- 3. Locate the group of editable annotations (rows) from a column (property) you want to batch edit (combine together) as a single annotation.
- 4. Shift click or Ctrl click to select (highlight) each annotation (row) entry. (Figure 343).

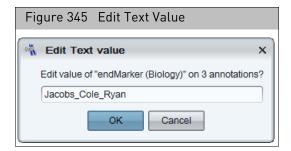


5. Right-click on a newly selected annotation. The following menu appears: (Figure 344)

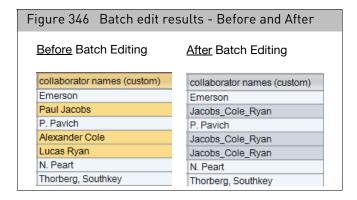


**IMPORTANT!** Only editable fields can be edited. If a field is non-editable, the Annotations Property Window pops-up. This window displays which fields are not editable (grayed out) versus those that are editable. Also, not all user-editable AED file fields may currently be edited from within the AED Editor. Some basic values (start, stop, type) cannot be edited in the AED Editor table directly. You MUST use "View/Edit Annotation Properties..." for editing the particular field in the annotation of interest.

- 6. Click Edit single property.
- 7. The following window appears: (Figure 345)



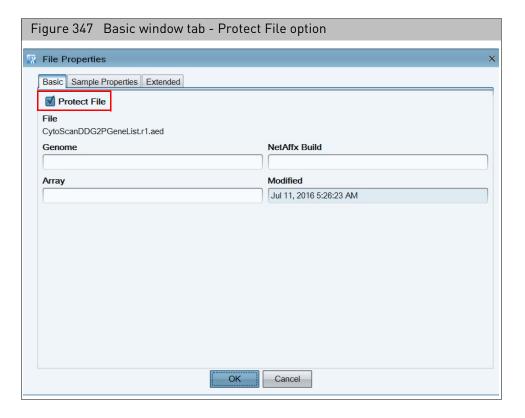
- 8. Enter an appropriate label to distinguish your new batch annotation entry, then click **OK**.
- 9. Your batch (group) of annotations appear as follows: (Figure 346)



### Protecting an AED file

Protecting an AED file provides a warning prior to adding an annotation to an AED file. This is an optional field and is intended to double-check the intention of adding an annotation to the selected AED file.

- 1. Right click on the AED file in the File Tree, select **View/Edit Properties**.
- 2. In the **Basic** tab window, click the **Protect File** check box, as shown in Figure 347.



When adding annotations to a Protected AED file, the following warning message appears. (Figure 348)



3. Click **Yes** to acknowledge the message, then click **OK** to add the annotation to the AED file.

Click **No** to return to the browser window without adding the annotation to the AED file.

**Note:** Protected AED files are noted with an asterisk, as shown in Figure 349.



### 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 segment data stored in AED files. For example, you could assign different colors to different categories:

- GainMosaic
- LossMosaic
- Gain
- Loss
- LOH

You can also perform comparisons for numerical values, coloring only values above or below a certain level.

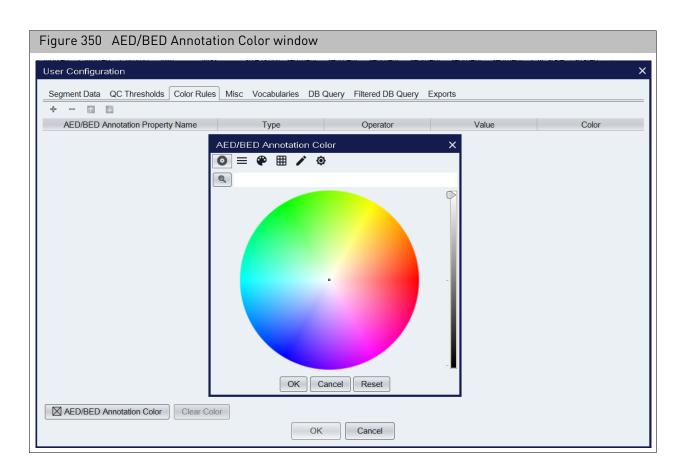
By default, ChAS displays the regions in an AED or BED file in a single color. Do one of the following to edit annotation color:

- 14
- Select a new default color for all AED or BED files (page 311).
- Create a color rule that specifies a color for annotations which meet user-specified requirements. See "Creating a color rule" on page 312.
- In the Detail View, choose a color for a particular annotation in the Annotation Properties window. See "Changing an annotation color" on page 317.

**Note:** An AED annotation color set in the Annotation Properties window (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.

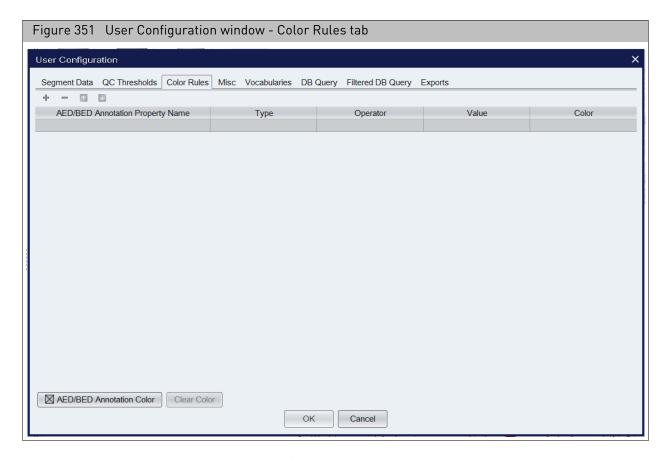
Selecting a new default color for loaded AED or BED files

- 1. Open the **User Configuration** window (click the button or select **Preferences** → **Edit User Configuration** on the menu bar).
- 2. In the Color Rules tab (Figure 350), click the AED/BED Annotation Color button.
- 3. In the window that appears, choose a color swatch or use the color controls to specify a color.
- 4. Click **OK** in the AED/BED Annotation Color window.



## Creating a color rule

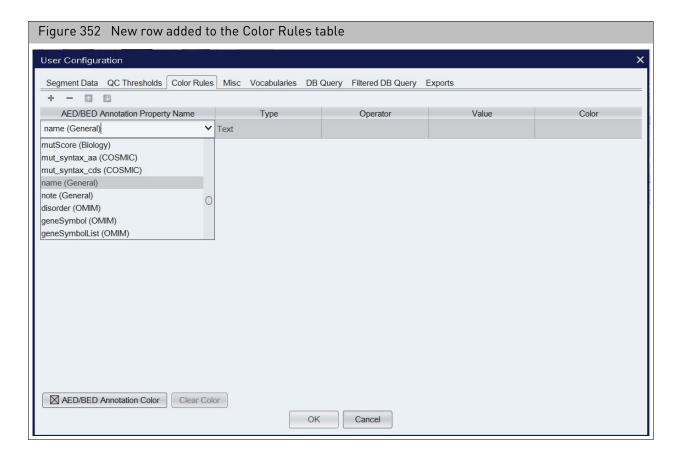
- 1. Click the button or select **Preferences** → **Edit User Configuration** on the menu bar.
- 2. The **User Configuration** window appears. (Figure 351)
- 3. Click the Color Rules tab.



The Color Rules tab has five columns:

Column	Description
AED/BED Annotation Property Name	Name assigned to the property.
Туре	Type assigned to the property (for example, text).
Operator	Type of comparison with value performed.
Value	Value assigned to the property.
Color	Color assigned to the region property and value.

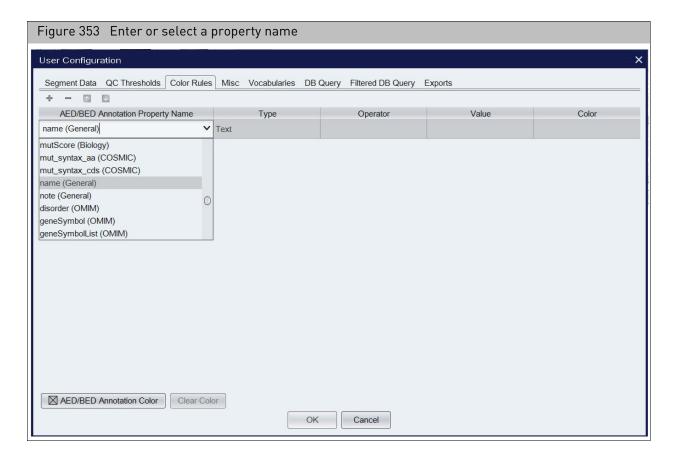
4. Click the **Add** + button at the top of the table (Figure 352).



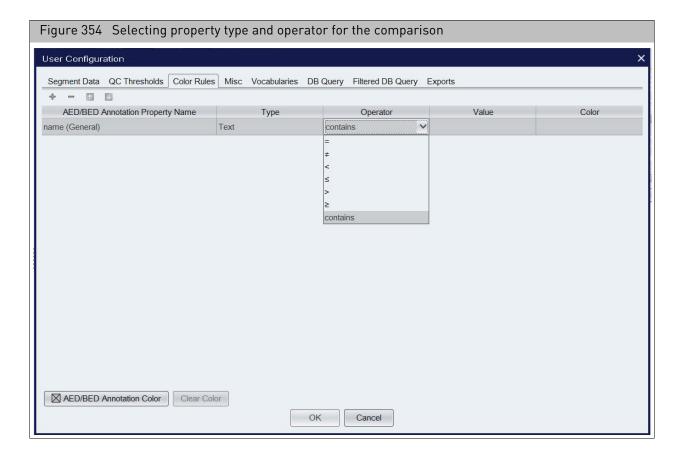
A new row appears in the table.

5. 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 (Figure 353). For more details, see Appendix C, "ChAS properties and types" on page 481.

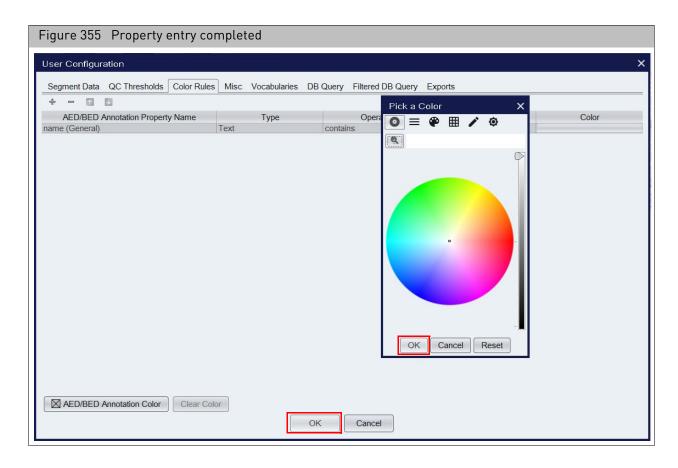
To delete a property, select the appropriate row, then click the **Remove** button.



- 6. Click in the row under **Type**, then select a property type from the drop-down list.
- 7. Click in the row under **Operator**, then select an appropriate operator from the drop-down list. (Figure 354)



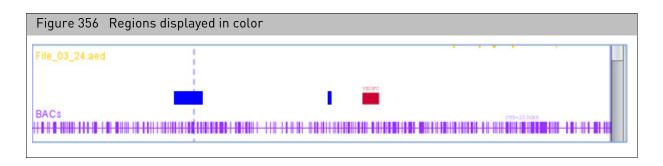
- 8. Click in the row under **Value**, then enter a value for the property.
- Click in the row under Color.
   The Pick a Color window appears. (Figure 355)
- 10. Select a color, then click **OK**.



Your property entry is completed.

11. Click **OK** to close the **Color Rules** window.

Regions that satisfy the property comparison are displayed in your selected color (for all views), as shown in Figure 356 on page 316.



## Changing an annotation color

1. Click Choose Color.

The **Pick a Color** window appears. (Figure 357)



- 2. Choose a color in the Swatches tab, or click the HSB or RGB tab to define a color.
- 3. Click OK.

**Note:** The color set in the Annotation Property window overrides the colors specified by a Color Rule created in the User Configuration window. For more details, see "AED/BED color rules" on page 310.

### 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).

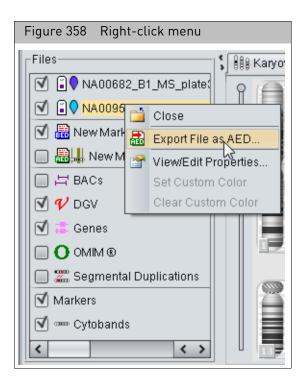
For more information on the AED file format, see page 471.

**Note:** AED or BED files created in ChAS 1.0, 1.0.1, or 1.1 do not automatically include the hg version.

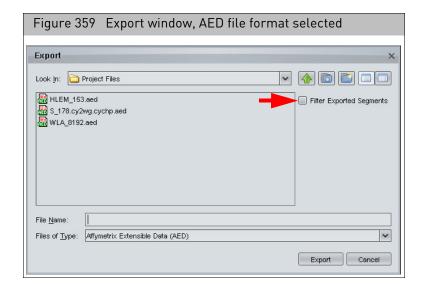
Position Data	Export to AED File	Export to BED File
Detected Segments for xxCHP files	Regions, names, and properties	Regions and names
Annotation Features in Reference Annotation files	Regions, names, and properties	Regions and names

Exporting position data as an AED or BED file

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 358)



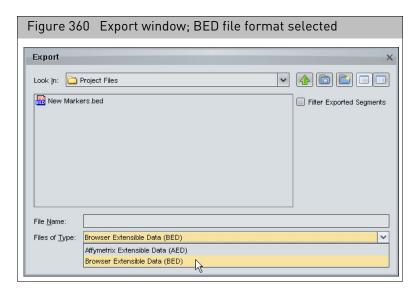
The **Export** window appears. (Figure 359)



Click the Filter Exported Segments check box 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 window pane.

- 3. Select a folder location for the file, as you normally would.
- To export to AED file format, enter a name for the file.
   To export to BED file format (Figure 360), enter a name for the file, then select Browser Extensible Data (BED) from the Files of Type drop-down list.



5. Optional: Select the Filter Exported Segments window to export only segments that meet filter criteria.

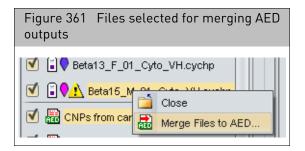
6. Click Export in the Export window.

The AED file is exported and can be loaded as a region information file or sent to another user.

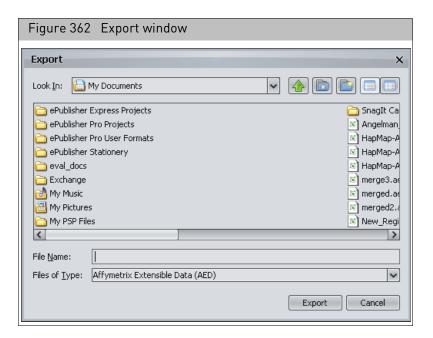
You can also merge feature position data from multiple different files.

#### Merging and exporting feature position information for multiple files

- 1. Select files in the File List by clicking on them while pressing the CTRL key.
- 2. Right-click the selected files, then click Merge Files to AED. (Figure 361)



The Export window appears. (Figure 362)



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. Also, 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.

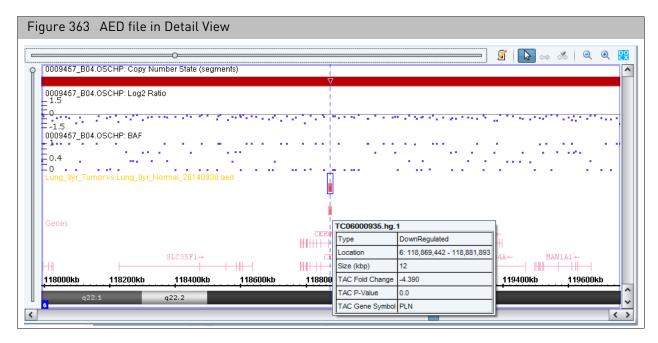
## Expression analysis AED file generation

AED files containing Gene Expression information and data can be created using Thermo Fisher Scientific Human Gene Expression arrays analyzed in the Transcription Analysis Console (TAC) 3.0 or higher.

For details on how to create an AED file containing Gene Expression/miRNA data, please see the Gene Expression Copy Number Analysis Quick Reference Card.

#### Viewing a Gene Expression AED file in ChAS

Load the AED file exported from TAC using File → Open.
 The AED is now visible in the Detail View. (Figure 363)



You can simultaneously view fold change from your Gene Expression dataset with copy number data from xxCHP files. Positive gene expression fold changes are represented by blue transcript cluster IDs. Negative fold changes are represented by red transcript cluster IDs. The deeper the color (blue or red) the larger the fold change.

### VCF files



### Loading VCF files

VCF files containing genotyping, copy number and structural variants can be visualized in the Detail View in conjunction with CHP file data.

VCF files are a flexible format, therefore you must use the following guidelines when formatting VCF files for use in ChAS.

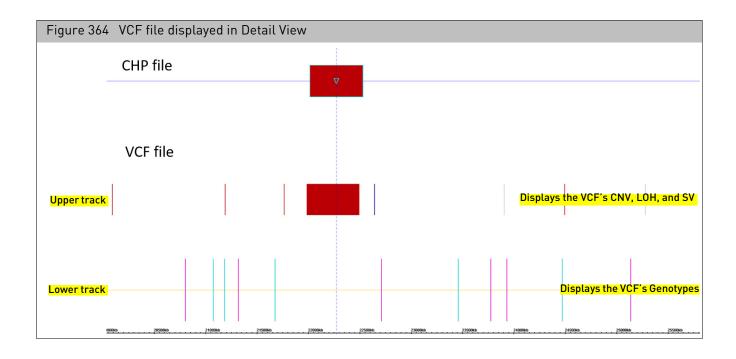
- Each VCF file must contain a single sample only.
- The VCF file will display genotype/indel/copy number/LOH and Structural Variation data.
- The VCF file must contain only 10 columns.
- The chromosome ID should use the format chr1, chr2 or 1, 2.
- Only chromosomes 1-22, X and Y are supported. Information from other sequences or variant chromosome assemblies will be ignored.
- The DATA column (column 10) can not be left blank/empty.
- A binary compressed VCF is not supported.
- The **vcf.gz** format can also be viewed in the ChAS browser.
- VCF entries representing structural variations (CNV, DEL, DUP, etc.) do not require GT values.
- Other VCF entries must contain **GT** in the FORMAT column, preferably in the first position.
- The length of structural variation segments should be indicated with a SVLEN value. If missing, the software auto-checks for LEN and END.

**IMPORTANT!** If your VCF file(s) do not strictly adhere to the above guidelines, the file(s) will not be compatible with, nor load into ChAS.

- 1. From your software, export your VCF file (based on the above guidelines).
- 2. Click **File**  $\rightarrow$  **Open**.
  - An Explorer window appears.
- 3. Navigate to your VCF file location, click to highlight it, then click **Open**.

A message appears indicating that the Genome Version could not be detected. If you are certain that the genome build for the VCF file matches the Genome Version loaded in the ChAS Browser, click **OK** to acknowledge the message.

The VCF file loads, then appears inside the Detail View. Copy number and structural variants are displayed in a upper track, while genotype calls are display on the lower track, as shown in Figure 364.



Genotype Data (Lower track)			
Data Type	Description	Color	
GT	Heterozygous Genotype	Cyan	
GT	Homozygous Alternate Genotype	Magenta	
GT	Homozygous Reference Genotypes	Gray	
Segment Data	Segment Data (Upper track)		
Data Type	Description	Color	
CNV	Alt allele = DUP	Blue	
CNV*	Copy Number Value >2 (>1 on Y)	Blue	
CNV	Alt allele = DEL	Red	
CNV*	Copy Number Value < 2 (<1 on Y)	Red	
LOH	Alt allele = LOH	Purple	
INS	Insertion	Gray	
INV	Inversion	Gray	
DUP:Tandem	Tandem Duplication	Gray	

<sup>\*</sup> CNV segments on the X and Y chromosomes may appear gray if the gender of the file cannot be determined which in turn, prevents the gain/loss status from being determined.

4. Optional: Mouse over a segment or a genotype in the VCF track to display more detail, as shown in the genotype example. (Figure 365)

Figure 365 VCF file mouse over details	
Type	VCF
Location	5: 180,374,484 - 180,374,485
Reference	С
Alternate	A
GT	0/1
Quality	233.25
Zygosity	Heterozygous
GT DNA	C/A

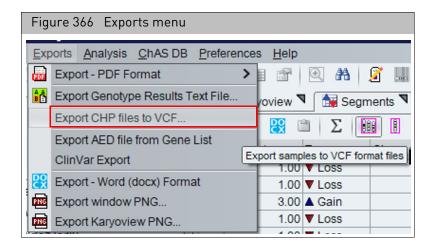
Mouse-ove	Mouse-over table definitions	
Type	File Type	
Location	Genome location for the SNP (or Indel)	
Reference	Reference Base	
Alternate	Variant base	
GT	Genotype Call, 0 = Reference Base, 1 = Variant Base	
Quality	Quality Score assigned from the assay platform	
Zygosity	Homozygous or Heterozygous Call	
GT DNA	Call in genotyping format (applies to segment data too)	

To link to TaqMan assays from the VCF file, see "Viewing and Ordering TaqMan assays for genotyping" on page 217. **Note:** Linking feature applies to hg38 analyses only.

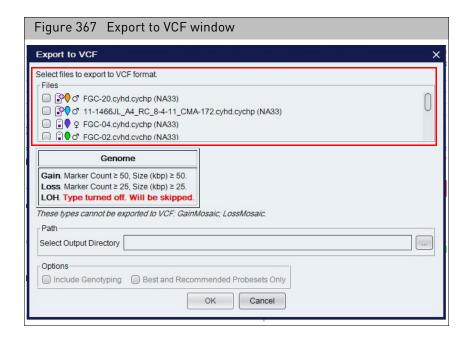
### **Exporting VCF files**

**IMPORTANT!** Before you can export VCF files, you must install RHAS on the same workstation.

Exporting CN and/ or variant data as a VCF file (for use in 3rd party browsers) 1. Click Exports  $\rightarrow$  Export CHP files to VCF. (Figure 366)



The Export to VCF window appears. (Figure 367)



- 2. Select the files to be exported in VCF format by clicking on the check box next to the file name.
- 3. Review the Genome filters to make sure the correct data will be exported into the VCF format. If not, click **Cancel**, update the filters accordingly, then repeat Step 1.

- 4. Click the Path **Browse** button.
  - A Explorer window appears.
- 5. Navigate to the location where you want to save your exported VCF, then click **Save**.
  - Your selected output directory appears.
- 6. Optional: Click the **Include Genotyping** check box to export the copy number segments and genotype data together. For HTCMA arrays, check **Best and Recommended Probesets Only** to export genotypes for those probesets the algorithm selected as the best and recommended.
- 7. Click OK.



# Displaying data in table views

#### This chapter includes:

- "Display overview"
- "Common table operations" on page 327
- "Sum, mean, and median calculator" on page 331
- "Saved table states" on page 332
- "Segments table" on page 337
- "Graphs table" on page 345
- "Variants table" on page 351
- "QC and sample info tab" on page 355
- "Chromosome summary data" on page 362
- "Auto-generated Autosome LOH percentage" on page 364
- "Searching results" on page 365
- "Finding intersections" on page 369

## Display overview

The data in the xxCHP files can be displayed and exported in tabular format, as well as the graphic representation in the Karyoview, Selected Chromosome View, and Detail View.

- "Selection details table" on page 210
- "CytoRegions table" on page 273
- "Viewing the overlap map table" on page 284

## Common table operations

The controls that are common to all tables are described in this section and include:

- "Standard tool bar controls" on page 328
- "Sorting by columns" on page 328
- "Changing the order of table columns" on page 329
- "Selecting columns to display or hide" on page 329
- "Sum, mean, and median calculator" on page 331

# Standard tool bar controls

The tool bar provides quick access to table options.

#### Left side



Export as TXT file. See "Exporting tables as TXT file" on page 423.



Export as PDF Report. See "Exporting table data into a PDF" on page 420.



Export as MS Word DOCX file. See "Exporting as Word (DOCX) format" on page 418.



Copy selected cells to clipboard. See "Exporting a segments table with modified segments to a TXT file" on page 425.



Calculate the sum, median, and mean 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.

#### Far Right



The number of rows in the table.



Opens the Select Columns window that enables you to choose the column headers to show or

The Export functions are described in "Exporting table data" on page 420.

#### 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.

#### Sorting a table by a single column

A triangle appears in the header.

1. Click in the header of the column to sort the table by that column's values.

- Triangle pointing upwards indicates ascending sort order.
- Triangle pointing downwards Min indicates descending sort order.
- 2. Click the header to toggle between Ascending, Descending, or no sort order.

**Note:** The **Type** column in the Segments table sorts segments based on the order that they appear in the **Data Types** window pane, not in an ascending/descending alphabetical order.

#### Performing a multi-column sort

- 1. Click in the header of the first column you want to sort on.
- 2. Click in the header of the second column.

3. 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 (Figure 368).

CytoRegion Type	CytoRegion	Chromosome	Min	Max	▲ Size (kbp)	▼ Segment ID		Segmen
Default User Annotation	CNP505	3	163,995,351	164,106,575		seq684	Beta10_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP2269	17	41,521,619			seg6392	Beta13_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP12506	17	26,040,482			seq6312	Beta13_F_01_Cyto_VH.cychp	▲ Gain
Default User Annotation	CNP2197	16	72,953,795			seg6151	Beta13_F_01_Cyto_VH.cychp	▲ Gair
Default User Annotation	CNP2156	16	21,422,575	21,498,841	76	seg5977	Beta13_F_01_Cyto_VH.cychp	▲ Gair
Default User Annotation	CNP147	1	194,997,658	195,068,695	71	seg587	Beta13_F_01_Cyto_VH.cychp	▲ Gair
Default User Annotation	CNP148	1	195,089,940	195,168,372	78	seg587	Beta13_F_01_Cyto_VH.cychp	▲ Gair
Default User Annotation	CNP933	6	32,539,530	32,681,749	142	seg572	Beta15_M_01_Cyto_VH.cychp	▼ Los:
Default User Annotation	CNP901	5	180,311,316	180,350,709	39	seg537	Beta15_M_01_Cyto_VH.cychp	▼ Los:
Default User Annotation	CNP12104	13	69,640,329	69,670,896	30	seg5218	Beta13_F_01_Cyto_VH.cychp	▼ Los:
Default User Annotation	CNP1952	13	68,149,981	68,166,243	16	seg5216	Beta13_F_01_Cyto_VH.cychp	▼ Los:
Default User Annotation	CNP12099	13	66,945,140	66,954,900	9	seg5214	Beta13_F_01_Cyto_VH.cychp	▼ Los:
Default User Annotation	CNP12094	13	65,094,528	65,103,708	9	seg5210	Beta13_F_01_Cyto_VH.cychp	▼ Los
Default User Annotation	CNP1946	13	63,122,789	63,134,693	11	seg5208	Beta13_F_01_Cyto_VH.cychp	▼ Los
Default User Annotation	CNP12093	13	64,198,454	64,209,619	11	seg5208	Beta13_F_01_Cyto_VH.cychp	▼ Los
Default User Annotation	CNP12091	13	63,227,094	63,303,323	76	seg5208	Beta13_F_01_Cyto_VH.cychp	▼ Los

# Changing the order of table columns

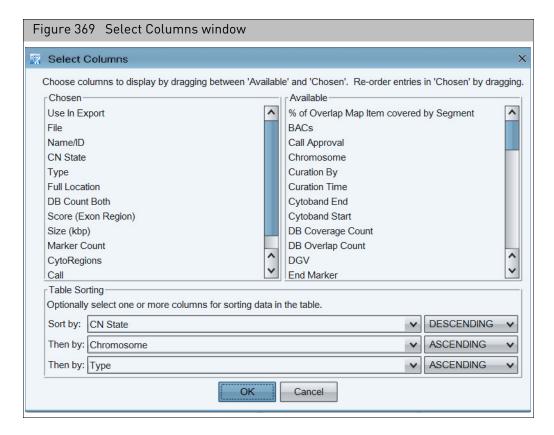
1. Click and drag in the column header to move the column to a new location.

# Selecting columns to display or hide

You can select columns using these two methods:

- Select Columns window.
- Right click on a column header, then choose **Select columns**.
- 1. Click the Select Columns tool bar button . The Select Columns window opens. (Figure 369)

**Note:** Specific items may vary, as they are dependent upon the type of table and data being displayed.



The columns in the left side pane are displayed in the table (in their default order).

- To hide columns within the table, move the column entry from the left (Chosen) pane into the right (Available) pane.
- To view different columns in the table, click and drag entries from the right pane to the left pane.
- Column order can be determined by clicking onto a column entry, then dragging it into its desired location (inside the left pane).
- Use the **Table Sorting** drop-down menus to sort your columns.

**Note:** These choices are auto-saved between sessions.

#### Using the rightclick menu

Right-click on a column header to perform the following:

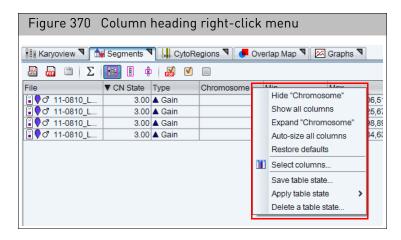
- Hide the selected column.
- Show all columns, including hidden ones.
- Expand selected column to display complete heading.
- Auto-size all columns.
- Restore the default selection of columns.
- Select other columns to hide or show.
- Create, select, save, and delete saved table states.

#### Using a column header's right-click menu

1. Right-click on the appropriate column header.

A menu appears. (Figure 370)

**Note:** Right-click menu items vary, as they are dependent upon the type of table and data that is displayed.



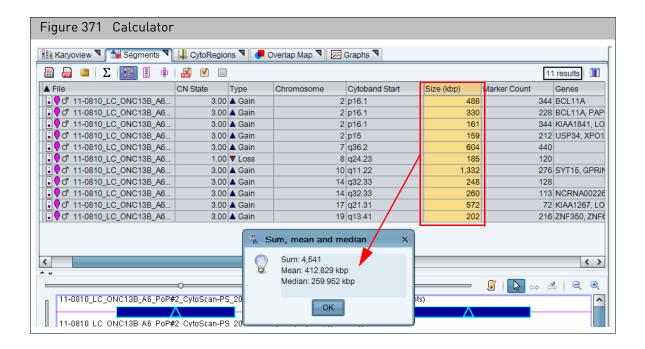
2. Click to select the option you want.

### Sum, mean, and median calculator

Use this calculator to calculate the sum, mean, and median of the selected or multiple numeric fields.

Calculating multiple numeric values in a column

- 1. Ctrl click or Shift click to highlight multiple numeric fields (MUST BE in the same column).
- Click \( \sum\_{\text{.}} \).
   Your multiple numeric values are calculated and summarized, as shown in. Figure 371.
- 3. Click OK



#### Saved table states

Saving multiple customized tables is a very useful tool, as it enables you to use a large number of columns in a table for interpreting the biology of a particular sample. On the other hand, saving a table state with a smaller subset of columns in a different order - may be more appropriate for including in a report.

In either case, having various Saved Table States enables you to switch back and forth between multiple sets of tables/columns without having to painstakingly recreate interpreting and reporting columns for each of your sample selections.

In the Segments Table, there are six pre-loaded Table States available for use in ChAS:

- Cytogenetic
- Default
- Oncology
- Oncomine Reporter
- ReproSeq
- ClinVar

There are two pre-loaded Table States available for use in the Variants table. See "Variants table" on page 351 for details.

- CytoScan HTCMA
- OncoScan

There are six pre-loaded Table States in the QC & Sample Info table based on relevant metrics for the array type loaded.

CytoScan QC View

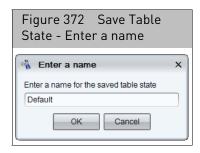
- Default
- CytoScan HTCMA QC View
- OncoScan QC View
- ReproSeq QC View
- SMN Sample Info View

**IMPORTANT!** You can restore your default settings at any time by right-clicking on a table column header and selecting Restore Defaults from the menu or selecting the Default Table State under Apply Table State.

# Saving the current segment table state to its default

- 1. Right-click on any column header.
  - A menu appears.
- 2. Click Save table state.

The following window appears: (Figure 372)

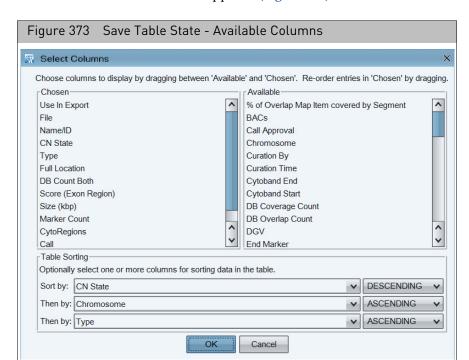


- 3. Type Default.
- 4. Click OK.

The table is now saved (as Default) to your User Profile for future use and/or reference.

# Adding columns to table states

1. Click the Select Columns icon. (Top right of Segments table)



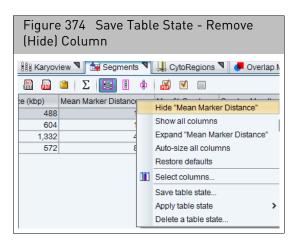
A list of available columns appears. (Figure 373)

- 2. Move the scroll bar downward to reveal more available columns.
- 3. Click and drag the column header inside the right pane into the desired location inside the left pane.
- 4. Click OK.

The additional column(s) is now added to your the left pane location. If the column is not in the position you want, click on the column, then drag and drop it into its correct spot.

Removing (hiding) columns in a table (for report use)

1. Right-click on a column header you want to remove (hide) from the table. The following menu appears: (Figure 374)



2. Click to select Hide "Column Name".

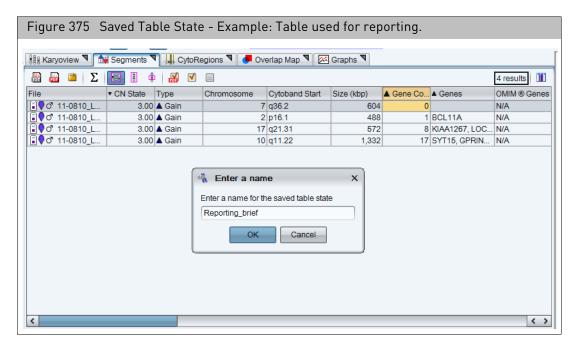
The column is now removed (hidden) from your table.

3. Repeat steps 1-2 to remove (hide) additional columns that are not required for a report.

# Saving a newly edited segment table state

- Right-click on any column header.
   A menu appears.
- 2. Click Save table state.

A pop-up window appears. (Figure 372)



- 3. Type a name for your new table state. (Example: **Reporting\_brief**)

  The saved table state remembers the columns that you selected, their order, their widths, and which ones were used for sorting.
- 4. Click OK.

The edited table is now saved to your User Profile for future use/reference.

# Applying previously saved table states

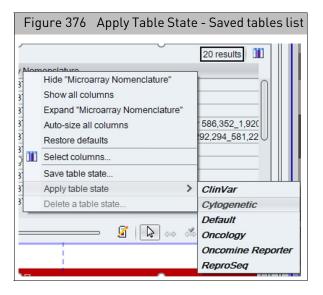
There are six Table States that are created automatically in ChAS.

- **Default**: Commonly used columns in any xxCHP file analysis
- Oncomine Reporter: Simplified Table State for export and use with Oncomine Reporter Software.
- Cytogenetic: Commonly used columns when analyzing constitutional samples.
- Oncology: Commonly used columns when analyzing cancer samples.
- ReproSeq: Standard columns for use with analyzing ReproSeq samples.
- ClinVar: All required columns for using the ClinVar export.
- 1. Right-click on any column header.

A menu appears.

#### 2. Click Apply table state.

A sub-menu appears displaying your saved tables, as shown in Figure 376.

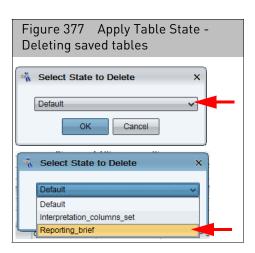


- 3. Click on the table you want to display. The table is now displayed.
- 4. (Optional) Repeat steps 1-3 to apply (open) other previously saved tables.

# Deleting previously saved table states

- 1. Right-click on any column header.
  - A menu appears.
- 2. Click Delete table state.

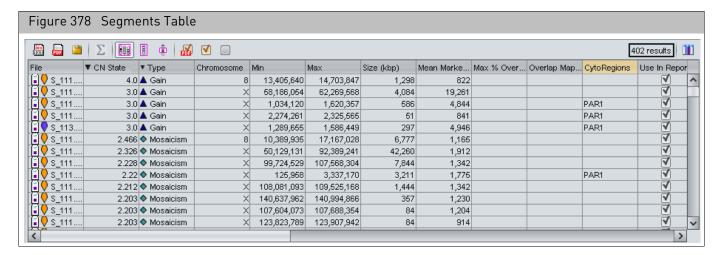
The following dialog window appears: (Figure 377)



- 3. Click the drop-down button to reveal your saved tables.
- 4. Click to select the table you want to delete.
- 5. Click OK
- 6. (Optional) Repeat steps 1-5 to delete other saved tables.

### Segments table

The Segments table (Figure 378) displays a list of the detected segments in the loaded sample data files.



Segment Table components:

- "Segments table tool bar" on page 338
- "Segments table" on page 338

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.

**Note:** Sorting the table based on the dynamically computed columns may be slow.

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.

# Segments table tool bar

The tool bar (Figure 379) provides quick access to table functions. The standard functions are described in "Standard tool bar controls" on page 328.



Use the **Use in Export** check boxes to <u>less</u> check/select all or <u>less</u> uncheck/deselect all segments displayed in the table.

**Note:** If the **Use in Export** column is hidden from the Segments table, then all rows are exported.

#### Segments table

In the Segments table, "N/A" can mean 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 Annotation file. "N/A" can also mean that a column which has been persisted to appear from a previous user profile, no longer has data in the current NetAffx Genomic Annotation file that is loaded. For instance the "Old OMIM" column will have "N/A" values when Browser annotations (version NA32.3 or higher) are loaded in ChAS 2.1 (or higher) software. See "ChAS browser NetAffx Genomic Annotations" on page 486.

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 can display each segment with the following information (the default set of columns in a new user profile may include only a subset of the total columns listed. For instructions on how to switch quickly between table column sets for a particular table, see "Saved table states" on page 332.

Materially Modified Segments (merged, created *de novo*, segments with edited start or end coordinates) and deleted segment have a different appearance in the Segments Table and export to TXT differently depending on the software settings. For more information, see "Exporting a segments table with modified segments to a TXT file" on page 425.

### Segment table columns

Prioritization segment prioritization in the segments table" on page 377.  Chromosome Chromosome on which the segment was found.  CN State 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 i entire length. This is due to the structure of the sex chromosomes. See "LOH segments on X ar Y chromosomes" on page 49 for more information.  Confidence (ReproSeq) The log likelihood that the called copy number state is not normal ploidy, example 2 on autosomic (reflects the likelihood of the region's ploidy number being different than the normal ploidy 2).  Curation By The current computer Operating System login ID and ChAS user profile name at the time that the Curation Time The time and date when the Curation field was last edited.  Cytoband End Cytoband in which the segment begins.  Cytoband Start Cytoband in which the segment begins.  CytoRegions Names of the CytoRegions with which the segment shares coordinates.  DB Both The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap The number of segments in the database meeting the Minimum Percent Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Setting up a ChAS DB query" on page 388.	Column	Description
Call Approval Cal User-editable field populated by a user-configurable drop list of Calls Call Approval Can be used as a bookmark for segments that have been reviewed. The Call term assigned based on Tier or Score Classification. For more information, see "Viewir segment prioritization in the segment stable" on page 377. Chromosome Chromosome on which the segment was found. CN State 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 i entire length. This is due to the structure of the sex chromosomes. See "LOH segments on X ar Y chromosomes" on page 49 for more information.  Confidence (ReproSeq) The log likelihood that the called copy number state is not normal ploidy, example 2 on autosom (reflects the likelihood of the region's ploidy number being different than the normal ploidy 2).  Curation By The current computer Operating System login ID and ChAS user profile name at the time that the Curation Time The time and date when the Curation field was last edited.  Cytoband End Cytoband in which the segment ends.  Cytoband Start Cytoband in which the segment begins.  CytoRegions Names of the CytoRegions with which the segment shares coordinates.  DB Both The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Overlap. This number can change depen	Item covered by	Overlap Map Item and the percentage by which it is covered by the segment.
Call Approval Call From The Call term assigned based on Tier or Score Classification. For more information, see "Viewin segment prioritization in the segments table" on page 377. Chromosome Chromosome Ochromosome on which the segment was found. CN State 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 i entire length. This is due to the structure of the sex chromosomes. See "LOH segments on X ar Y chromosomes" on page 49 for more information.  Confidence (ReproSeq) Ireflects the likelihood of the region's ploidy number being different than the normal ploidy 2!.  Curation By The current computer Operating System login ID and ChAS user profile name at the time that the Curation field was last edited.  Curation Time The time and date when the Curation field was last edited.  Cytoband End Cytoband in which the segment ends.  Cytoband Start CytoRegions Names of the CytoRegions with which the segment shares coordinates.  DB Both The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Court Both. See "Querying a segment from the segment table" on page 390.  DB Coverage The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Overlap. This number change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	BACs	List of BACs that share coordinates with the segment.
The Call term assigned based on Tier or Score Classification. For more information, see "Viewing segment prioritization in the segments table" on page 377.  Chromosome  Chromosome on which the segment was found.  CN State  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 in entire length. This is due to the structure of the sex chromosomes. See "LOH segments on X ard Y chromosomes" on page 49 for more information.  Confidence [ReproSeq]  The log likelihood that the called copy number state is not normal ploidy, example 2 on autosome (reflects the likelihood of the region's ploidy number being different than the normal ploidy 2).  Curation By  The current computer Operating System login ID and ChAS user profile name at the time that the Curation field was last edited.  Curation Time  The time and date when the Curation field was last edited.  Cytoband End  Cytoband in which the segment ends.  Cytoband Start  Cytoband in which the segment begins.  CytoRegions  Names of the CytoRegions with which the segment shares coordinates.  DB Both  The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  The number of segments in the database meeting the Minimum Percent Overlap. This number change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click	Call	User-editable field populated by a user-configurable drop list of Calls
Prioritization segment prioritization in the segments table" on page 377.  Chromosome Chromosome on which the segment was found.  CN State 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 i entire length. This is due to the structure of the sex chromosomes. See "LOH segments on X ar Y chromosomes" on page 49 for more information.  Confidence (ReproSeq) The log likelihood that the called copy number state is not normal ploidy, example 2 on autosome (reflects the likelihood of the region's ploidy number being different than the normal ploidy 2).  Curation By The current computer Operating System login ID and ChAS user profile name at the time that the Curation Time The time and date when the Curation field was last edited.  Cytoband End Cytoband in which the segment ends.  Cytoband Start Cytoband in which the segment begins.  CytoRegions Names of the CytoRegions with which the segment shares coordinates.  DB Both The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Setting up a ChAS DB query" on page 388.	Call Approval	Can be used as a bookmark for segments that have been reviewed.
CN State  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 i entire length. This is due to the structure of the sex chromosomes. See "LOH segments on X ar Y chromosomes" on page 49 for more information.  Confidence (ReproSeq)  The log likelihood that the called copy number state is not normal ploidy, example 2 on autosome (reflects the likelihood of the region's ploidy number being different than the normal ploidy 2).  Curation By  The current computer Operating System login ID and ChAS user profile name at the time that the Curation field was last edited.  Curation Time  The time and date when the Curation field was last edited.  Cytoband End  Cytoband in which the segment ends.  Cytoband Start  Cytoband in which the segment begins.  CytoRegions  Names of the CytoRegions with which the segment shares coordinates.  DB Both  The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  The number of segments in the database meeting the Minimum Percent Overlap. This number cot change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.		The Call term assigned based on Tier or Score Classification. For more information, see "Viewing segment prioritization in the segments table" on page 377.
The expected Copy Number State on the X chromosome in normal males is not constant over i entire length. This is due to the structure of the sex chromosomes. See "LOH segments on X ar Y chromosomes" on page 49 for more information.  Confidence (ReproSeq) The log likelihood that the called copy number state is not normal ploidy, example 2 on autosome (reflects the likelihood of the region's ploidy number being different than the normal ploidy 2).  Curation By The current computer Operating System login ID and ChAS user profile name at the time that the Curation field was last edited.  Cytoband End Cytoband in which the segment ends.  Cytoband Start Cytoband in which the segment begins.  CytoRegions Names of the CytoRegions with which the segment shares coordinates.  DB Both The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Count Both. See "Querying a segment from the segment table" on page 390.  DB Coverage The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap The number of segments in the database meeting the Minimum Percent Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	Chromosome	Chromosome on which the segment was found.
Curation By	CN State	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 "LOH segments on X and
Curation Time The time and date when the Curation field was last edited.  Cytoband End Cytoband in which the segment ends.  Cytoband Start Cytoband in which the segment begins.  CytoRegions Names of the CytoRegions with which the segment shares coordinates.  DB Both The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Count Both. See "Querying a segment from the segment table" on page 390.  DB Coverage The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap The number of segments in the database meeting the Minimum Percent Overlap. This number of change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.		The log likelihood that the called copy number state is not normal ploidy, example 2 on autosomes (reflects the likelihood of the region's ploidy number being different than the normal ploidy 2).
Cytoband End Cytoband in which the segment ends.  Cytoband Start Cytoband in which the segment begins.  CytoRegions Names of the CytoRegions with which the segment shares coordinates.  The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Count Both. See "Querying a segment from the segment table" on page 390.  DB Coverage The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap The number of segments in the database meeting the Minimum Percent Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	Curation By	The current computer Operating System login ID and ChAS user profile name at the time that the Curation field was last edited.
Cytoband Start  CytoBand in which the segment begins.  CytoRegions  Names of the CytoRegions with which the segment shares coordinates.  The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Count Both. See "Querying a segment from the segment table" on page 390.  DB Coverage  The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	Curation Time	The time and date when the Curation field was last edited.
CytoRegions  Names of the CytoRegions with which the segment shares coordinates.  The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Count Both. See "Querying a segment from the segment table" on page 390.  DB Coverage  The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	Cytoband End	Cytoband in which the segment ends.
DB Both  The number of segments in the database meeting BOTH the Minimum Percent Overlap and the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Count Both. See "Querying a segment from the segment table" on page 390.  DB Coverage  The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	Cytoband Start	Cytoband in which the segment begins.
Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Count Both. See "Querying a segment from the segment table" on page 390.  DB Coverage  The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	CytoRegions	Names of the CytoRegions with which the segment shares coordinates.
can change depending on whether the "match only same gain/loss type" box is checked. Right click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Settin up a ChAS DB query" on page 388.  DB Overlap  The number of segments in the database meeting the Minimum Percent Overlap. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	DB Both	same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the
change depending on whether the "match only same gain/loss type" box is checked. Right-clic on the a row in the Segment Table. From the menu, click DB Overlap Count. See "Setting up a ChAS DB query" on page 388.	DB Coverage	The number of segments in the database meeting the Minimum Percent Coverage. This number can change depending on whether the "match only same gain/loss type" box is checked. Right-click on the a row in the Segment Table. From the menu, click DB Coverage Count. See "Setting up a ChAS DB query" on page 388.
DCV List of DCV variations that share coordinates with the comment	DB Overlap	
List of Dov variations that share coordinates with the segment.	DGV	List of DGV variations that share coordinates with the segment.
DGV-GS List of curated Database of Genomic Variants considered "Gold Standard" that share coordinate with the segment.	DGV-GS	List of curated Database of Genomic Variants considered "Gold Standard" that share coordinates with the segment.
End Marker The array marker name which marks the end of the segment.	End Marker	The array marker name which marks the end of the segment.

Column	Description	
Evidence	Provides information on which annotations the segment overlapped. For more information, see "Viewing segment prioritization in the segments table" on page 377.	
File	File the segment was detected in.	
Filtered DB Both	The number of segments in the database meeting the Minimum Percent Overlap and Minimum Percent Coverage and the selected Filter Criteria.	
Filtered DB Coverage	The number of segments in the database meeting the Minimum Percent Coverage and the selected Filter Criteria.	
Filtered DB Overlap	The number of segments in the database meeting the Minimum Percent Overlap and the selected Filter Criteria.	
FISH Clones	List of FISH clones that share coordinates with the segment.	
Following Marker	The array marker just below the segment in the data track used as input for the segment.  Note: This column is only applicable to CNState Gain and Loss segments.	
Following Marker Location	The coordinate location of the array marker just below the segment in the data track used as input for the segment.	
	Note: This column is only applicable to CNState Gain and Loss segments.	
Full Location	Chromosome Start and Stop in a user-friendly format for use in external databases.	
Gene Count	A count of the gene names listed in the Genes column	
Genes	List of RefSeq genes from the Genes track that share coordinates with the segment. Identically named gene isoforms are NOT repeated.	
Inheritance	User-editable field populated by a user-configurable drop list of Inheritance.	
Marker Count	Number of markers in the segment.	
Materially Modified By	The current computer Operating System login ID and ChAS user profile name at the time that the segment was last materially modified.	
Materially Modified Segment	Indication that segment was previously merged, deleted, or had its start or end boundary, type, or state altered by a ChAS user. (ChAS-based processes of Smoothing and Joining are not "Modifications", nor are making Calls or Interpretations, in this context).	
Materially Modified Time	The time and date when the Segment was last materially modified.	
Max	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 Appendix E, "Genomic position coordinates" on page 487).	
Max % Coverage	The highest percentage by which a segment covers some item(s) in the Overlap Map.	
Max % Overlap	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".	
Mean Log2 Ratio	The mean of all the Log2 Ratio values contained in the segment.	

Column	Description		
Mean Marker Distance	Length of the segment in base pairs divided by the number of markers in the segment.		
Mean Weighted Log2 Ratio	The mean of all the Weighted Log2 Ratio values contained in the segment.		
Median Log2 Ratio	The median of all the Log2 Ratio values contained in the segment.		
Microarray Nomenclature	An ISCN-based description of the segment.		
Microarray Nomenclature ISCN 2013	ISCN 2013 based description of the segment.		
Min	Start position of segment.		
Number of Overlap Map Items	Number of Overlap Map items which share genomic coordinates with the segment.		
Old OMIM	The column formerly known as "OMIM" in the ChAS software when NetAffx Genomic Annotation Browser files version NA32.1 and earlier are loaded. This column's information is now out of date and has been superseded by the newer OMIM Genes and OMIM Phenotype Loci columns present in NA32.3 and above browser annotation files.		
OMIM Gene Count	A count of the OMIM Gene names listed in the OMIM Genes column.		
OMIM Genes	List of OMIM Genes that share coordinates with the segment.		
OMIM Phenotype Loci	List of OMIM Phenotype Loci that share coordinates with the segment.		
OMIM Region Phenotype	Lists the Gene Title of any region that overlaps with a copy number or LOH segment.		
Oncomine Report	A drop-down list of annotations compatible with the Oncomine Reporter Software. For details on this application, go to: https://www.thermofisher.com/order/catalog/product/A34298		
Overlap Map Items (% of Segment overlapped)	Item(s) in the Overlap Map which overlap the segment, followed by the percentage by which the segment is overlapped by that Item.		
Preceding Marker	The array marker just above the segment in the data track used as input for the segment.  Note: This column is only applicable to CNState Gain and Loss segments.		
Preceding Marker Location	The coordinate location of the array marker just above the segment in the data track used as input for the segment.  Note: This column is only applicable to CNState Gain and Loss segments.		
Precision (ReproSeq)	The log likelihood that the called copy number state is different than next closest copy number state (reflects the likelihood that the precise ploidy number is correct).		
Protein Coding Genes	List of protein coding RefSeq genes from the Genes track that share coordinates with the segment. Identically named gene isoforms are NOT repeated.		
Protein Coding Genes Count	Number of Protein Coding RefSeq genes that shapre coordinates with the segment.		

Column	Description
Protein Coding Ensembl Genes	List of protein coding Ensembl Genes annotations that share coordinates with the segment. Identically named gene isoforms are NOT repeated.
Protein Coding Ensembl Genes Count	Number of Protein Coding Ensembl Genes that share coordinates with the segment.
Sample UUID	Unique identifier for the CHP file.
Segment Interpretation	User-editable field for free-text interpretation on the segment
Segment Label	A label comprised of the segment's Type, State, and Filename.
Segment Name/ID	File-specific identifier assigned to the detected segment.
Segmental Duplications	List of Segmental Duplications that share coordinates with the segment.
Size (kbp)	Size of the segment.
Smoothed/Joined/ XON merged	Indication that the segment was created by smoothing, joining or merging two or more segments in the initial segment detection.
sno/miRNA	List of sno/miRNA features that share coordinates with the segment.
Start Marker	The array marker name which marks the beginning of the segment.
Summarized Log 2 Ratio	The median of the LR, after transformation to adjust for individual marker responsiveness.
Tier or Score	The assigned Tier or Score value based on the Segment Prioritization method selected. When using Tier based, the column will display the assigned Tier. When using Score based, the column will display the score value based on the annotations the segment overlaps. For more information, see "Viewing segment prioritization in the segments table" on page 377.
Туре	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 window pane.
Use in Report	Allows manual selection of Segments for export to a Segments Table PDF, DOCX, or Text rather than all segments in the table.
XON Region Level	The annotation Level assigned to this region of the genome.

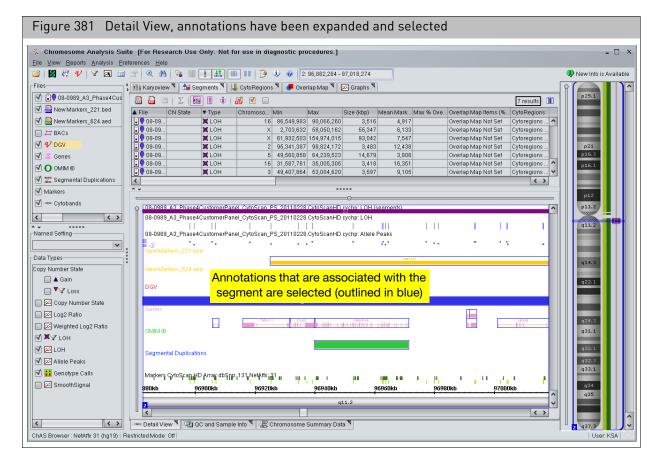
Obtaining all annotations associated with a segment

 Select a segment in the Segments table, Karyoview, Detail View, or Selected Chromosome View, then press Ctrl+Space or right-click on the segment to select Zoom to selection from the menu.

The exact length of the selected segment fills the entire width of the Detail View (Figure 380).

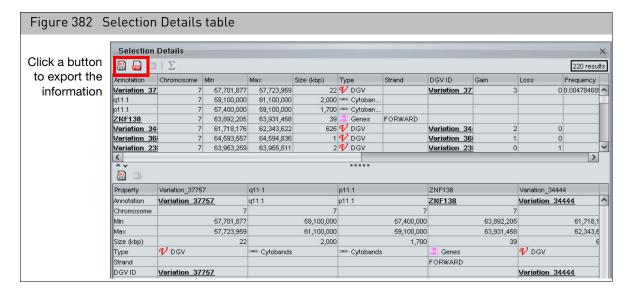


2. Click the **!!!** tool bar button to expand all annotation tracks.



- 3. 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. (Figure 381)
- 4. Click the tool bar button.

The Selection Details table appears (Figure 382). It includes all of the items selected in the Detail View. For more details on the table, see "Selection details table" on page 210.



For details on exporting from the Segment table, see "Exporting table data" on page 420.

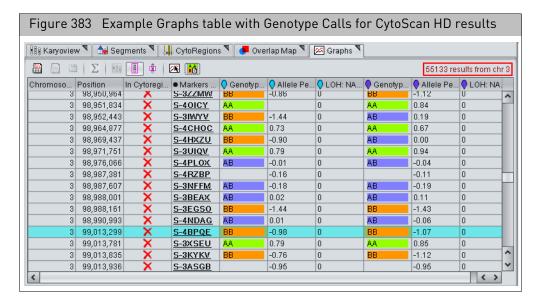
### Graphs table

**IMPORTANT!** The results from ChAS are for <u>Research Use Only</u>. Not for use in diagnostic procedures.

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. 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 array results (Figure 383).



The table displays each data point in the displayed graphs.

To highlight markers in the views or the table:

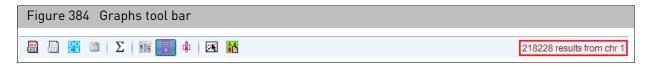
- Click in a row of the table to place the cursor on the marker in the Chromosome and Detail Views.
- Click on a marker in the Selected Chromosome or Detail View to highlight the marker in the Graphs table.

# Graphs table properties

The Graphs columns display:

Column	Description
In CytoRegion	Whether marker is located in a CytoRegion or not:
	✓ In CytoRegion
	Not in CytoRegion.
	See "Using CytoRegions" on page 268 for more information.
Markers	Marker ID. Right-click to link to NetAffx information about the marker.
	<b>Note:</b> For efficiency reasons, it is not possible to sort the table on this column.
Genotype	The SNP genotype call.
Graph Data types	● Log2Ratio: NA18526_C09_01.cychp
	The Column header displays:
	Color Nib
	Data Type
	Name of sample file
	The table cell displays the value for the marker.

#### Graph Table Tool bar



The tool bar (Figure 384) enables you to:

- Export data in TXT format only. See "Exporting tables as TXT file" on page 423
- View data for chromosome and selected chromosome region only. You cannot display data from the whole genome in Graphs tab. You can export this data for the whole genome using the "Displaying and exporting data from the analysis workflow" on page 93.
- Export genotype results

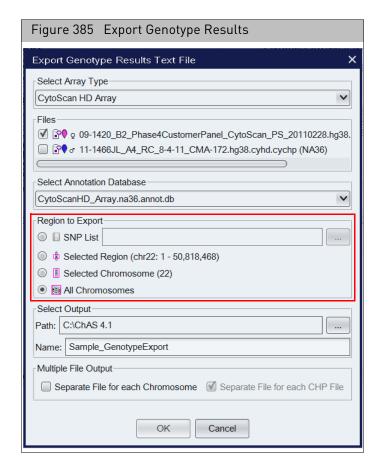
The Graphs Settings button opens the Graph Settings panel, enabling you to change the style of graph, scale, and other features for the data graphs. See "Changing graph appearance" on page 198.

# To export genotype calls

**Note:** Exporting genotypes is not available for OncoScan or ReproSeq data.

1. Click the tool bar button. Alternatively, select **Reports** → **Export Genotype Results Text File** from the menu bar.

2. In the window that appears (Figure 385 on page 348), select the array type, results file(s) (CYCHP), and annotation database to use for the export.



- 3. Specify a region to export:
  - **SNP List**: SNPs specified in a user-created SNP list (.txt).
  - Selected Region: SNPs included in the chromosome region selected in the Karyoview.
  - **Selected Chromosomes**: 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, as shown in Figure 386 on page 349.

The export will not proceed without one of these column headers.

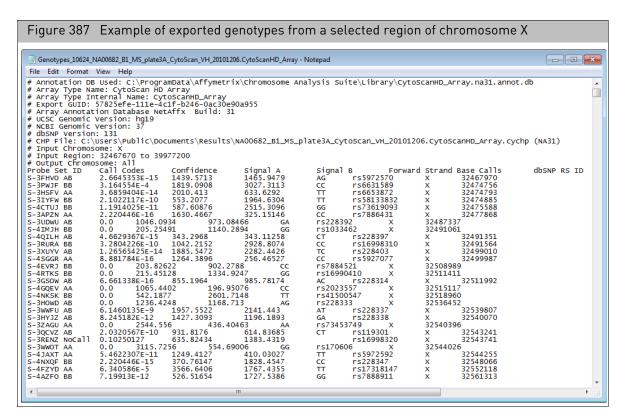
	Figure 386 Example SNP list		
	Α		
1	Probe Set ID		
2	S-3FHVO		
3	S-3PWJF		
4	S-3HSFV		
5	S-3IYFW		
6	S-4CTUJ		
7	S-3APZN		
8	S-3UDWU		

- 4. Enter the path name or click the **Browse** button to select a folder for the output.
- 5. 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.
- 6. 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
None	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.
Separate File for each Chromosome	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.
Separate File for each CYCHP File	Creates one text file per CYCHP file. Each file contains genotype calls for all chromosomes.
Separate File for each Chromosome and Separate File for each CYCHP File	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.

#### 7. Click OK.

**Note:** Exporting of Genotypes may take several minutes to complete, as this process is dependent on the total number of SNPs selected for export.



The exported text file (Figure 387) includes information about the analysis (for example, array type, NetAffx annotation database, hg version, and chromosome).

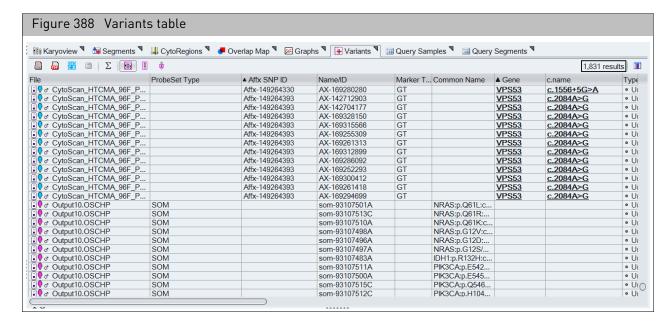
**Note:** If the option "Separate File for each CYCHP File" was not selected, many of the headers will be repeated for each CYCHP file. The header titles will be appended with the CYCHP file name to indicate which file the column belongs to.

The column headers report the following information:

Column	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.

#### Variants table

For OncoScan FFPE and CytoScan HTCMA arrays only. The Variants table (Figure 388) displays the somatic mutation information from OncoScan FFPE arrays and/or the variant information from CytoScan HTCMA arrays.



The Variants table components include:

- Tool bar (below)
- Variants table on page 351

To highlight segments in the views or in the table:

- Click the row of the table to zoom to the somatic mutation in the Karyoview, Selected Chromosome View and Detail View.
- Click on the somatic mutation in the Karyoview, Selected Chromosome View or Detail View to highlight the somatic mutation in the Somatic Mutations Table.
  - Detected: Large, bright green circles in the graphic views denote somatic mutations of high confidence (OncoScan) or variants called het, hom, NoCall, NRP (anything that isn't a major homozygous call)
  - Undetected: Small, dark gray dots in the graphic views denote somatic mutations and variants in which the wild type or major homozygous genotype was called.

**Note:** Calls can be removed from the table (and graph view) by de-selecting the Undetected or Detected check box(es). in the Data Types window.

#### Tool bar

The Tool bar (Figure 389) provides quick access to table functions. Its standard functions are described in "Standard tool bar controls" on page 328.



#### Variants table

The table can display each mutation with the following information (the default set of columns in a new user profile may include only a subset of the total columns listed below).

#### CytoScan HTCMA

Column	Description
Affx SNP ID	Unique Thermo Fisher Scientific generated identifier for the SNP.
Alt Allele	The call for the first alternate allele associated with a non-normal phenotype.
Alt Code	The Call Code (A,B,C,D,) of the Alternate Allele.
Alternate Name	Displays the alternative names for the variant.
Associated Phenotype	Displays the Phenotype that is associated with the variant.
c.name	Displays standard variant nomenclature based on coding DNA reference sequences.
Chromosome	The chromosomal location of the variant.
File	Name of the sample file.
g.name	Displays the genomic coordinates for the variant.
Gene	Name of heritable genetic sequence that encodes proteins.
Inheritance Pattern	Method of Inheritance (Example: AR (Autosomal Recessive), XLR (X-linked Recessive).
Marker Type	The type of marker (Indel, SNP, CN).
Max	Ending genomic position for the variant.
Min	Starting genomic position for the variant.
Name/ID	Thermo Fisher Scientific identifier for the probeset.
p.name	Displays the change in protein translation for the variant.
Recommended Probeset	A quality control metric determined by SNP Polisher algorithm that chooses the best probesets querying a SNP.
Ref Allele	The call for the reference allele associated with a normal phenotype.
Ref Code	The Call Code (A,B,C) of the Reference Allele.

Column	Description
RSID	dbSNP ID
Size	Size in bp of the variant.
Transcript	RefSeq transcript ID associated with the c.name.
Туре	Undetected/Detected. Detected are those mutations with any call other than the wild-type or major homozygous genotype.
Variant Status	Status of the variant based on the genotype (i.e. Not Detected, Het, Hom, NoCall, NRP).
Variant Status Alt Allele	Severity status for the variant mapped to Alt Allele.
Variation ID	ClinVar ID

### OncoScan CNV Plus

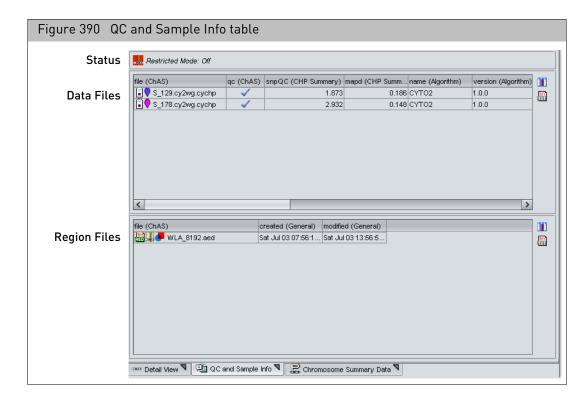
Column	Description
accession number	RefSeq Gene accession number.
Channel	CEL file from which the signal is measured. "A" is the AT CEL, "C" is the GC CEL.
Chromosome	Chromosome on which the somatic mutation is found.
Common Name	Abbreviated description of the mutations to which this ProbeSet is known to respond. The name has the form [Gene]:[amino acid change for mutation]:[cDNA change for mutation]. In the event that the ProbeSet cannot differentiate among multiple mutations to which it can respond, the slash (/) delimits the multiple known mutations.
COSMIC ID	The identifier of the mutation as listed in the COSMIC database, which is a catalogue of somatic mutations in cancer. More information on these mutations can be found at: http://cancer.sanger.ac.uk
Event	Describes if the probeset is a point mutation, deletion, insertion or sequence variant.
Event Type	Describes if the event is missense, frame-shift, in-frame insertion or deletion.
File	Name of the OSCHP file the somatic mutation is in.
Fwd 3' flank	Sequence flanking the mutation at the 3' end.
Fwd 5' flank	Sequence flanking the mutation at the 5' end.
Genes	RefSeq gene that shares coordinates with the somatic mutation.
High Threshold	High confidence MutScore threshold. Measurements equal to or greater than this threshold are called "High confidence," describing the likelihood that the mutation is present.
Low Threshold	Lower confidence MutScore threshold. Measurements with a MutScore below this value are called "Undetected". Measurements equal to or greater than this threshold but less than the High Threshold are called "Lower confidence," describing the likelihood that the mutation is present.
Max	Stop position of the somatic mutation.

Column	Description			
Min	Start position of the somatic mutation.			
Mutation (aa)	Wild type > mutant amino acid change on the coding strand.			
Mutation (nt)	Wild type > mutant nucleotide change on the coding strand.			
Mutation Syntax (aa)	Encoding of which nucleotide was changed and its location in the CDS.			
Mutation Syntax (cds)	Encoding of which amino acid was changed and the location of the codon.			
MutScore	Measures somatic mutation probeset response. The stronger the response, the more likely it is that the somatic mutation is present. The MutScore calculation depends on the algorithm version. The newer MutScore calculation also corrects for sample-specific effects, and thereby reduces false positive calls, which were sample specific.  For algorithm versions 1.0 - 1.2 (ChAS 3.0 and earlier, OncoScan Console 1.2 and earlier): MutScore.old = (measured quantile normalized signal - median signal for this marker in the reference model file) / (95th percentile signal for this marker in the reference model file) - median signal for this marker in the reference model file).  For algorithm versions 1.3 and newer (ChAS 3.1 and newer, releases of OncoScan Console after 1.2):  MutScore.new = (MutScore.old - median MutScore.old for this sample) / standard deviation of MutScore.old for this sample (where standard deviation is calculated for all but the num-out-std strongest MutScore.old for this sample, median is calculated for all but the num-out-med strongest MutScore.old for this sample, and the used median is the maximum of zero and the measured median).			
Name/ID	Thermo Fisher Scientific identifier for the marker.			
ProbeSet Type	Probeset Type is Somatic Mutation.			
Size (bp)	The size of the Somatic mutation in base pairs.			
Source DB	Cosmic database version.			
Strand	Coding strand of the associated gene (Forward or Reverse).			
Туре	HighConfidence, LowerConfidence or Undetected call as to the presence of the mutation.			

**Note:** Changes made to an OSCHP file in the Somatic Mutation Viewer Application requires the OSCHP file to be reloaded into the ChAS Browser to reflect the change made to the sample.

### QC and sample info tab

The QC and Sample Info tab (Figure 390) in the lower pane displays information about the loaded Data and Region files.



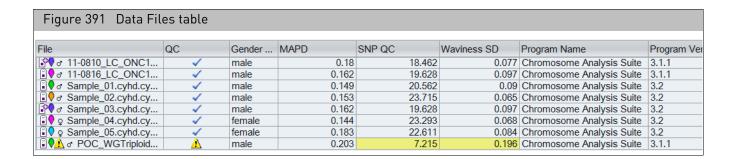
The top section displays Status for Restricted Mode (see "Using restricted mode" on page 276)

The tables display information on:

- Loaded data files. "QC and sample information table" on page 355
- Loaded region files. "Loaded AED/BED files table" on page 361

# QC and sample information table

The QC table has six pre-loaded Table States allowing you to quickly toggle to the relevant information based on array type. For detailed information, see "Saved table states" on page 332.



The six pre-loaded QC Table States and their column descriptions are shown below:

### CytoScan QC view

Column	Description			
Antigenomic Ratio	Ratio of median intensity antigenomic control probes/median intensity all copy number probes.			
File	File Name			
Genome Version	Build of the genome (i.e. hg19, hg38)			
MAPD	Median Absolute Pairwise Difference value. See Appendix G for detailed information.			
Median Raw Intensity	Pre-processed Median signal of the array.			
QC	In or Out of QC bounds.			
Sex	Gender call for the sample. See "Gender call algorithms" on page 360).			
SNPQC	SNP QC value. Median Absolute Pairwise Difference value. See Appendix G for detailed information.			
Waviness SD	A global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation. See Appendix G for detailed information.			
Waviness Segment Count	Number of raw segments without any post-processing.			

#### Default QC view

Column	Description			
Algorithm Name	Name of the algorithm used in processing the array.			
Algorithm Version	Version of the algorithm used in processing the array.			
Annotation File	Name of the Annotation file used to create the xxCHP file.			
Array Type	Type of array used in the analysis.			
Autosome LOH	The proportion of LOH on chromosome 1 to 22.			
Cel Pair Check	Inspects each pair of intensity (*.cel) files to determine whether the files have been properly paired and assigned to the correct channel.			
Created	Date the xxCHP file was created.			
File	File Name			

Column	Description			
Low Diploid flag	An essential part of the algorithm is the identification of "normal diploid" markers in the cancer samples. This is particularly important in highly aberrated samples. The normal diploid markers are used to calibrate the signals so that "normal diploid markers" result i a log2 ratio of 0 (e.g. copy number 2). The algorithm might later determine that the "normal diploid" markers identified really correspond to (for example) CN=4. In this case the log2 ratio gets readjusted and TuScan ploidy will report 4. Occasionally (in about 2% of samples the algorithm cannot identify a sufficient number of "normal diploid" markers and no "normal diploid calibration occurs. This event triggers "low diploid flag" = YES. In this case the user needs to carefully examine the log2 ratios and verify if re-centering is necessary.			
MAPD	Median Absolute Pairwise Difference value. See Appendix G for detailed information.			
Modified	Date the xxCHP file was last modified.			
ndSNPQC	QC metric for SNP probes that is derived from polymorphic SNP probes in normal diploid regions.			
ndwavinessSD	Measure of variation of probes in normal diploid regions that are insensitive to short-rang variation and focus on long-range variation.			
Parameter File	Name of the chasparam file used to create the xxCHP file.			
QC	In or Out of QC bounds.			
Reference	Reference Model file used in the single sample analysis.			
Sex	Gender call for the sample. See "Gender call algorithms" on page 360).			
SNPQC	SNP QC value. Median Absolute Pairwise Difference value. See Appendix G for detailed information.			
SNPQC Type	ND or non-ND			
TuScan %AC	If % AC = 100%, we return "homogeneous" because it could be 100% normal or 100% tumor. If % AC =NA, the percent aberrant cells could not be determined and TuScan returns non-integer CN calls. This metric is an algorithmically determined estimate of the % of aberrant cells in the sample.			
TuScan Ploidy	The most likely ploidy state of the tumor before additional aberrations occurred. TuScan Ploidy is assigned the median CN state of all markers, provided that %AC could be determined and integer copy numbers are returned. If %AC cannot be determined, NA (No Available) is reported for both ploidy and %AC.			
Waviness SD	A global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation. See Appendix G for detailed information.			

### CytoScan HTCMA QC view

Column	Description		
DishQC (DCQ)	Measures the amount of overlap between two homozygous peaks created by non polymorphic probes. DQC of 1 is no overlap, which is good. DQC of 0 is complete overlap, which is bad.		
File	File Name		

Column	Description			
MAPD	Median Absolute Pairwise Difference value. See Appendix G for detailed information.			
QC	In or Out of QC bounds.			
QC Call Rate	Percentage of autosomal SNPs with a call other than NoCall (measured at the Sample QC step).			
QC Het Rate	Percentage of SNPs called AB (i.e. the heterozygosity) for autosomal SNPs (measured at the Sample QC step).			
Sex	Gender call for the sample. See "Gender call algorithms" on page 360).			
SMN MAPD	Median Absolute Pairwise Difference value calculated from the CNVMix algorithm for the SMN pipeline.			
SMN WavinessSD	A global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation from the CNVMix algorithm for the SMN pipeline.			
SNPQC	SNP QC value. Median Absolute Pairwise Difference value. See Appendix G for detailed information.			
Waviness SD	A global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation. See Appendix G for detailed information.			

### OncoScan QC view

Column	Description			
Cel Pair Check	Inspects each pair of intensity (*.cel) files to determine whether the files have been properly paired and assigned to the correct channel.			
Cel Pair Check Concordance	Percentage of SNPs that match between the AT and GC arrays.			
File	File Name			
Genome Version	Build of the genome (i.e. hg19, hg38)			
Low Diploid flag	An essential part of the algorithm is the identification of "normal diploid" markers in the cancer samples. This is particularly important in highly aberrated samples. The normal diploid markers are used to calibrate the signals so that "normal diploid markers" result in a log2 ratio of 0 (e.g. copy number 2). The algorithm might later determine that the "normal diploid" markers identified really correspond to (for example) CN=4. In this case the log2 ratio gets readjusted and TuScan ploidy will report 4. Occasionally (in about 2% of samples) the algorithm cannot identify a sufficient number of "normal diploid" markers and no "normal diploid calibration occurs. This event triggers "low diploid flag" = YES. In this case the user needs to carefully examine the log2 ratios and verify if re-centering is necessary.			
MAPD	Median Absolute Pairwise Difference value. See Appendix G for detailed information.			
ndSNPQC	QC metric for SNP probes that is derived from polymorphic SNP probes in normal diploid regions.			
ndWaviness SD	Measure of variation of probes in normal diploid regions that are insensitive to short-range variation and focus on long-range variation.			

Column	Description		
Number nd	Number of probes called normal diploid by the algorithm.		
Percentage nd	Percentage of probes called normal diploid by the algorithm		
QC	In or Out of QC bounds.		
Sex	Gender call for the sample. See "Gender call algorithms" on page 360).		
TuScan %AC	If % AC = 100%, we return "homogeneous" because it could be 100% normal or 100% tumor. If % AC = NA, the percent aberrant cells could not be determined and TuScan returns non-integer CN calls. This metric is an algorithmically determined estimate of the % of aberrant cells in the sample.		
TuScan Log 2 Ratio adjustment	Log 2 ratio determined from TuScan algorithm needed to "center" the diploid region of the sample (around Log $2 = 0$ ).		

### ReproSeq QC view

Column	Description			
Application (ReproSeq)	Name of Application from the ReproSeq assay.			
Batch File (ReproSeq)	lame of the batch file downloaded from Ion Reporter.			
Chr MA Ratio	Ratio of Mitochondrial/Autosome.			
File	File Name			
MAPD	Median Absolute Pairwise Difference value. See Appendix G for detailed information.			
Sex	Gender call for the sample. See "Gender call algorithms" on page 360).			
Single File (ReproSeq)	Name of the current file from a Batch File downloaded from Ion Reporter.			
Workflow (ReproSeq)	Name of the workflow run in Ion Reporter.			

### SMN Sample Info view

Column	Description		
DishQC (DCQ)	Measures the amount of overlap between two homozygous peaks created by non polymorphic probes. DQC of 1 is no overlap, which is good. DQC of 0 is complete overlap, which is bad.		
File	File Name		
MAPD	Median Absolute Pairwise Difference value. See Appendix G for detailed information.		
QC	In or Out of QC Bounds.		
QC Call Rate	Percentage of autosomal SNPs with a call other than NoCall (measured at the Sample QC step).		
QC Het Rate	Percentage of SNPs called AB (i.e. the heterozygosity) for autosomal SNPs (measured at the Sample QC step).		
Sex	Gender call for the sample. See "Gender call algorithms" on page 360).		

Column	Description			
SMN MAPD	Median Absolute Pairwise Difference value calculated from the CNVMix algorithm for the SMN pipeline.			
SMN WavinessSD	A global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation from the CNVMix algorithm for the SMN pipeline.			
SMN1 (variant)	Genotype call for SMN variant(s) as defined in the SMN.SNP.list For more details, see the RHAS User Guide.			
SMN1 CN	Copy number for SMN1.			
SMN1 Exon7	Copy number for Exon 7 of SMN1.			
SMN1 Exon8	Copy number for Exon 8 of SMN1.			
SMN2 CN	Copy number for SMN2.			

Other data from the header of the Sample Data file or the ARR file can also be selected for display in the Select Columns window.

You can only hide or display columns by using the Column Select window, at the right of the table.

**Note:** or samples run through the Normal Diploid Analysis for CytoScan, the ndSNPQC and ndwavinessSD metrics can be viewed in the QC Information Tab, but will not flag a sample as pass/fail.

# Gender call algorithms

The table below explains which algorithm is used to make the gender calls for the different arrays.

The CytoScan Array uses 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.

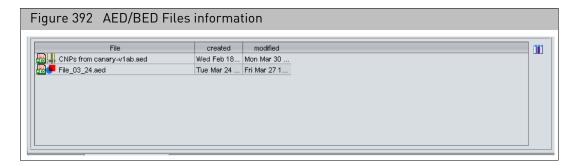
**Note:** For more details how the array-specific algorithms call LOH segments for the X or Y chromosome, see "LOH segments on X and Y chromosomes" on page 49.

Software/Array Type	Gender Call Algorithm	Call	Gender Call Confidence
ChAS 1.X, 2.X and 3.X/4.X CytoScan Arrays	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

Software/Array Type	Gender Call Algorithm	Call	Gender Call Confidence
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

# Loaded AED/BED files table

This table (Figure 392) displays information about the loaded Region (AED or BED) files.



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 window, at the right of the table.

### Chromosome summary data

The Chromosome Summary Data tab has two components:

- Chromosome Summary Data
- Auto-Generated Autosome LOH Percentage

The Chromosome Summary Data table (Figure 393) summarizes particular data across each chromosome in the loaded sample data files.

The available data types are:

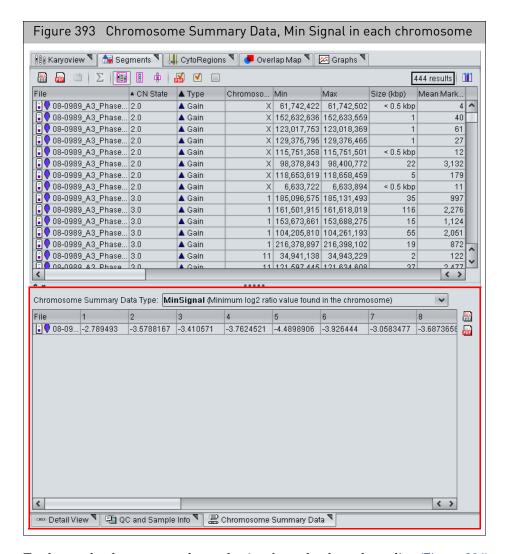
**Note:** ReproSeq Anueploidy data is not supported within this tab.

- Min Signal minimum Log2 Ratio value found in the chromosome
- Median Signal median Log2 Ratio value found in the chromosome
- Max Signal maximum Log2 Ratio value found in the chromosome
- Median CN State median calibrated Log2 Ratio
- Mosaicism median mosaicism mixture value

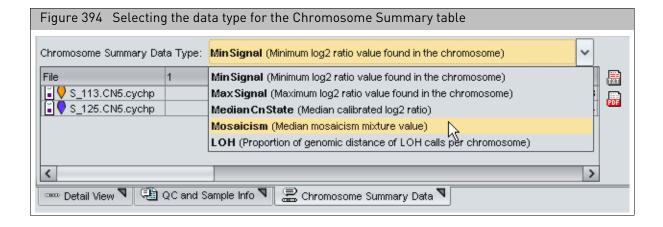
**Note:** The mosaicism CN state value is not an integer due to cell populations with different CN state values. In the Chromosome Summary Data table, the mosaicism value indicates how much the median CN state value is above or below two, the normal CN state value for autosomes and X in females. For 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



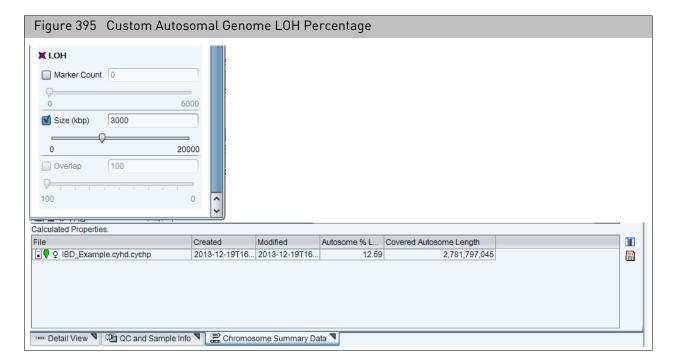
To choose the data type, make a selection from the drop-down list. (Figure 394)



#### Auto-generated Autosome LOH percentage

The percentage LOH displayed is calculated for the Autosome based on the filter size set for LOH in the Filters Window. (Figure 395)

The Covered Autosome Length is the base pairs of the Autosome covered by probes.



IMPORTANT! You must check the LOH Segment Data type to view the sample's percent LOH.

### Searching results

The Search function enables you to search:

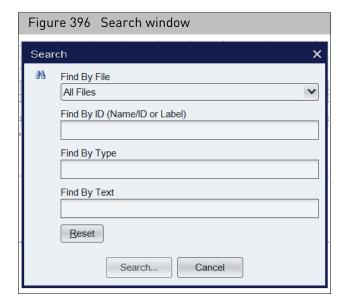
- Detected Segments
- Reference Annotations
- Loaded Region Information Files

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
- 1. From the View menu, select **Search** or click the upper tool bar's icon. The Search window opens. (Figure 396)



2. Search all files in the File List.

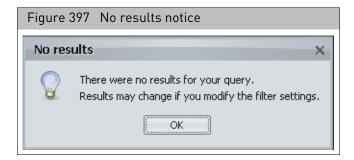
Alternatively, click the All Files drop-down to select the file you want to search.

- 3. In the **Find By ID** text field, enter the ID/Name you want to search. A wildcard (\*) is not required when performing searches, however the "\*" can be used to narrow searches. For example, performing the search *HOX* returns SHOX and RHOXF1 findings, while using a "\*" (\*HOX), returns SHOX.
- 4. Optional: Use the **Find By Type** and/or **Find By Text** fields to enter a type name (not case sensitive) for the search.

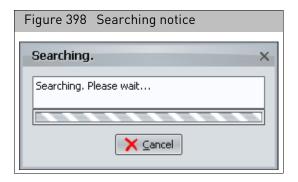
You can enter:

- Names for types of reference annotation features (Genes, DGV, etc.)
- Names for types of segments (Loss, Gain, etc.)
- 5. Click **Search**... to begin the search.

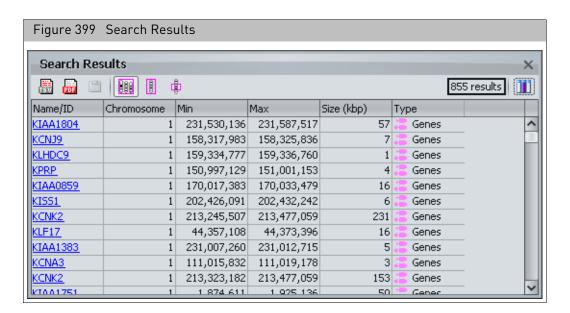
If no results are found, the following notice appears. (Figure 397)



If the search takes more than a few seconds, a **Searching...** window appears. (Figure 398)

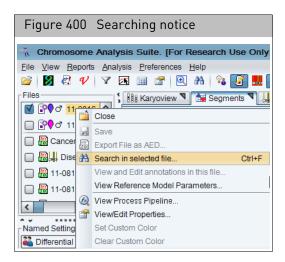


If results are found, the Search Results table opens. (Figure 399)

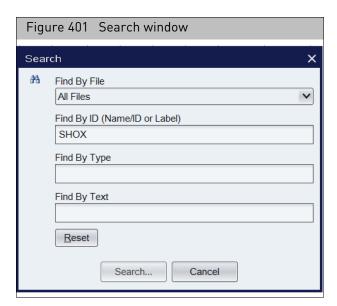


# Searching within a selected file

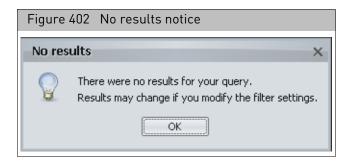
1. Right-click on a selected (checked) file inside the **Files** list pane.



- 2. Click to select **Search in selected file...** (Figure 400) The Search window opens. (Figure 401)
- 3. Click the appropriate radio button(s), enter your search criteria, then click **Search**.



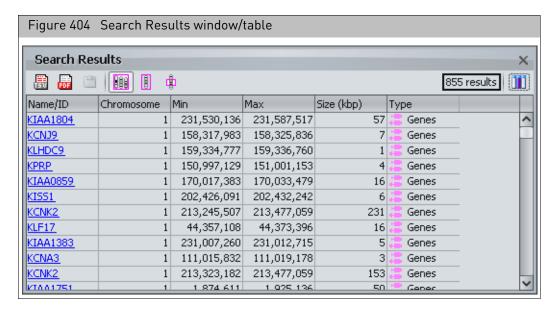
If no results are found, the following notice appears. (Figure 402)



If the search takes more than a few seconds, an In Progress notice appears. (Figure 403)



If results are found, the **Search Results** window/table opens. (Figure 404)



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" on page 327).

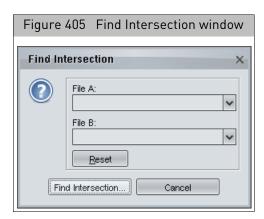
The Search Results table displays the following information:

Column	Description
Chromosome	Chromosome the items are located in.
Label	Name or ID of the item.
Max	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 Appendix E, "Genomic position coordinates" on page 487).
Min	Starting position of the item.
Size (kbp)	Size of the item.
Туре	Type of item.

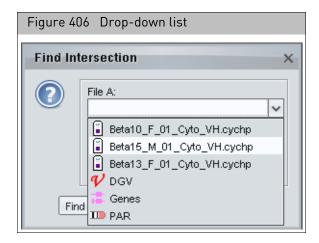
## Finding intersections

The **Find Intersection** feature enables you to find segments and regions that overlap for different:

- Detected Segments
- Reference Annotations
- Loaded Region Information Files
- From the View menu, select Find Intersections...
   The Find Intersection window opens. (Figure 405)



2. Select the first file for the comparison from the File A drop-down list. (Figure 406)



The list shows the available Sample files, Region Information Files, and Reference Annotations.

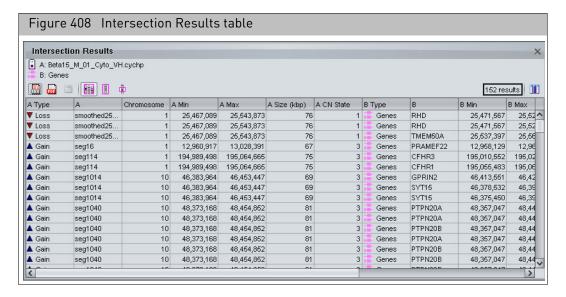
**Note:** Only files that are check marked in the Files List appear in the Match File drop-down list.

- 3. Select the second file from the File B drop-down list.
- 4. Click Find Intersection...

The Finding Intersection notice opens. (Figure 407)



When the comparison is finished, the Intersection Results table opens. (Figure 408)



The table displays the names of the A and B files above the tool bar.

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" on page 327).

#### The Intersection Results table displays the following information:

Column	Description
% A Touching B	How much of the A item is covered by the B item.
% B Touching A	How much of the B item is covered by the A item.
Α	Identifier used for item in A file.
A CN State	Copy number of the segment in file A.
A Max	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 Appendix E, "Genomic position coordinates" on page 487).
A Min	Starting position of the A item.
A Size (kbp)	Size of the A item.
А Туре	Type of item in A file with overlap.
В	Identifier used for item in B file.
B CN State	Copy number of the segment in file B.
В Мах	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 Appendix E, "Genomic position coordinates" on page 487)
B Min	Starting position of the B item.
B Size (kpb)	Size of the B item.
В Туре	Type of Item in B file with overlap.
Chromosome	Chromosome the items are located in.
Shared Size (kbp)	Size of the overlap.



# **Prioritizing segments**

Segment Prioritization enables the sorting of copy number segments based on user-defined relevance with overlapping annotations.

There are two segment prioritization options:

- "Tier-based prioritization"
- "Score-based prioritization" on page 378

**Note:** The following NetAffx Genomic Annotation files are required for full use of all the parameter in Segment Prioritization. Please use files: NetAffxGenomicAnnotations.Homo\_sapiens.hg19.na20200828/ NetAffxGenomicAnnotations.Homo\_sapiens.hg38.na20200828 (or more current) for optimal segment prioritization results.

**IMPORTANT!** All data results from the Segment Prioritization process should be manually reviewed.

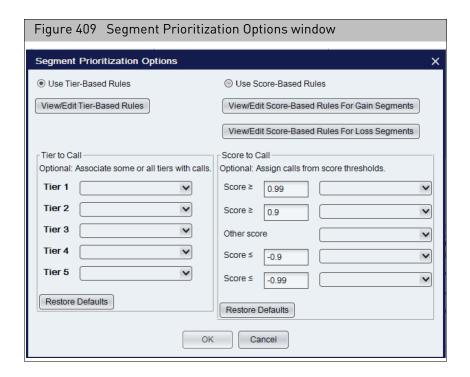
## Tier-based prioritization

The tier-based prioritization assigns Tiers 1-5 to copy number segments that overlap defined annotations. Copy number segments that overlap an annotation with an assigned tier will get have that tier assignment. If a copy number segment is assigned multiple tiers, the lowest tier number will be assigned to the copy number segment. The exception to this general rule are annotations for DGV-GS and ChAS DB Count both. If a segment is meets the criteria for either of these annotation, the tier assignment will take priority over lower number tiers.

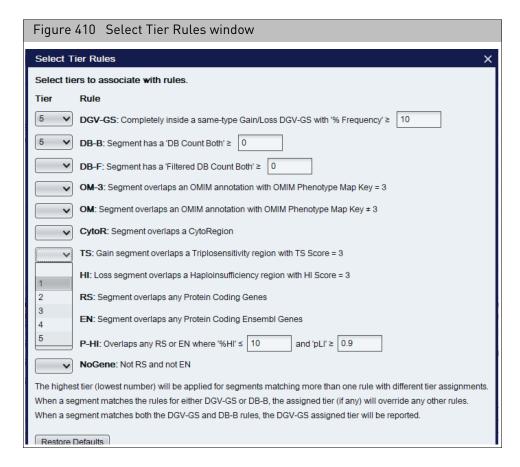
# Configuring the tier-based option

1. From the Segment Table tool bar, click the button.

The Segment Prioritization Options window appears. (Figure 409)



Click the View/Edit Tier-Based Rules button.
 The Select Tier Rules window appears. (Figure 410)



3. To prioritize your copy number segments in a tier-based order, assign a tier value to the rule(s) you want to use. To do this, click the drop-down arrow adjacent to the rule, then click the tier value (1-5) you want to assign to it. A Tier 1 assignment denotes the highest rule priority, while Tier 5 is the lowest.

**Note:** Not every Rule requires an assigned tier. See the table below for Rule definitions.

**IMPORTANT!** For copy number segments that meet rule(s) with different tier assignments, the highest tier (lowest number) will be assigned to the segment. **Example:** A copy number segment overlaps a rule assigned as Tier 1 and also a rule that is Tier 3, the segment will be assigned as a Tier 1 since that is the higher of the 2 tiers. Two Rules are exceptions: DGV-GS and DB-B, if either of these rules are met, the assigned tier overrides any other tiers. If both of these rules are met, the tier assignment for DGV-GS is assigned.

Tier-based rule/evidence	Description
DGV-GS	Database of Genomic Variants - Gold Standard. The copy number segment is completely contained within an entry of like type (Gain or Loss) from the Gold Standard DGV track and meeting a defined frequency. Default frequency is >=1%. This rule WILL override higher ranked tiers based on the tier selected in this rule.
DB-B	ChAS DB Both. Compare the copy number segment to the ChAS DB Both column data. If the number of entries in this column exceeds the defined threshold, then the copy number segment will be assigned the tier associated with this rule (unless the DGV-GS rule is also met). This rule WILL override rules with higher ranked tiers (with the exception of DGV-GS).
DB-F	ChAS DB - Filtered. Compare the copy number segment to the Filtered ChAS DB Both column. Example: Set the Filtered ChAS DB query to filter on segments in the database with the Call 'unknown significance'. If the copy number segment overlaps enough segments in your ChAS DB called 'unknown significance'', then the selected Tier will be assigned.
OM-3	Any OMIM Genes annotation that is dark green in color. Dark Green is assigned for phenotype map key 3 OMIM records indicating the molecular basis is known; a mutation has been found in the gene
ОМ	Any OMIM Genes annotation that is NOT dark green in color. See OM-3 (above).
Cyto-R	CytoRegions file. The segment overlaps any region in the assigned CytoRegion file. For more information on CytoRegions, see "Using CytoRegions" on page 268.
TS	Triplosensitivity. The copy number segment overlaps an entry in the Triplosensitivity track which has an assigned TS_score of 3.
HI	Haploinsufficiency. The copy number segment overlaps an entry in the Haploinsufficiency track which has an assigned HI_score of 3.
RS	RefSeq. The copy number segment overlaps an entry in the Protein Coding Genes Track.
EN	User-editable field populated by a user-configurable drop list of Calls
P-HI	The copy number segment overlaps a Protein Coding Gene or Protein Coding Ensembl Gene with predicted haploinsufficiency values meeting the defined thresholds. pLI derived from gnomAD (https://gnomad.broadinstitute.org/) and %HI derived from DECIPHER (https://decipher.sanger.ac.uk/).
NoGene	The copy number segment does not overlap any known Protein Coding Gene or Protein Coding Ensembl.

4. Once all desired rules have an assigned tier, click **OK** to save the selections and return to the Segment Prioritization Options window.

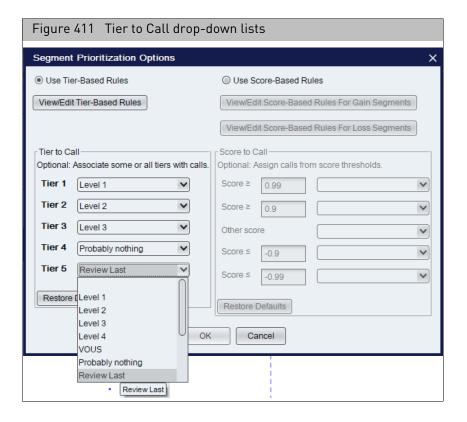
Click **Cancel** to return to the Segment Prioritization Options window without saving any tier assignment changes.

Click **Restore Defaults** to return to the installation settings.

#### Tier to call settings

You can assign a Call to represent each tier. The contents of the drop-down list was generated from the Calls Vocabulary list. There are a set of default 'Calls", but this list can be customized, as detailed in "Using the calls feature" on page 254.

1. Click on the drop-down list adjacent to the Tier(s) you want to assign a Call to, then click on a selection, as show in Figure 411.



**Note:** Tiers are not required to have a Call assigned. Unassigned Tiers will appear blank in the Segment Table's **Calls From Prioritization** column.

2. After your Tier to Call assignments are complete, click **OK**.

# Viewing segment prioritization in the segments table

Three new segment prioritization columns now appear in the Segment Table, as shown in Figure 412.



- **Call From Prioritization**: Displays the Call associated with the Tier assigned to the copy number segment.
- Evidence: Displays the abbreviation representing the rules met based on which annotations the copy number segment overlaps. See the Tier-based rule/evidence table of definitions above for more details.

**Note:** A copy number segment that does not overlap any rules with an assigned tier will display *No rules match*.

• **Tier or Score**: This will be a number from 1-5 representing the Tier that was assigned to the copy number segment based on the user-defined Tier-Based rules selected.

If the Call from Prioritization assignments are correct, they can be accepted as the Calls for each segment.

#### To do this:

1. Click the **button**.

The Call from Prioritization assignments will be copied into blank cells in the Call column. **Note:** Any Calls manually assigned will remain in the Call column and not be overwritten. Segments hidden by the filters will not have calls copied from the Calls from Prioritization column.

A confirmation message appears (Figure 413) summarizing the Call from Prioritization assignments are to be copied into the Call column.



2. Click **Yes** to acknowledge the message.

**Note:** Any Call can be manually adjusted by clicking on the Call cell and selecting a new option from the drop-down list, as detailed in "Using the calls feature" on page 254.

### Score-based prioritization

The Score-Based prioritization is a research-based adaptation that is similar, but not identical to, the guidelines proposed in the Riggs et al. 2019 paper: *Genetics in MEDICINE*, Published online: 06 November 2019.

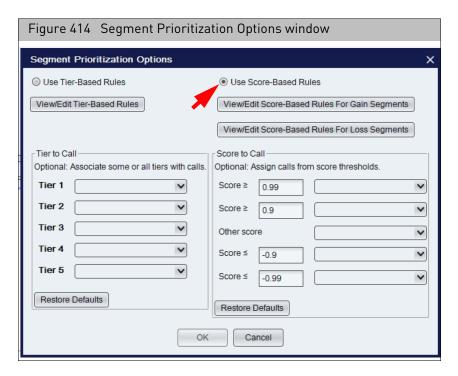
This segment prioritization method assigns numeric values (i.e. scores) based on the overlap of a copy number segment with public and/or private annotations. Default score assignments are based on the aforementioned paper.

- A copy number segment's final score is summed based on the rules the segment matches.
- The score assigned to a copy number segment is then associated with a call based on user defined thresholds.
- Overlap between copy number segments and annotations are performed on all transcripts for a given gene that are >=90% of the size of the gene coordinates in the Triplosensitivty and Haploinsufficiency tracks and have identical gene symbols.
- Segment prioritization applies to the following segment types: Gains, Mosaic Gains, XON Region Gains, Loss, Mosaic Loss, and XON Regions Loss.

# Configuring the score-based option

1. From the Segment Table tool bar, click the button.

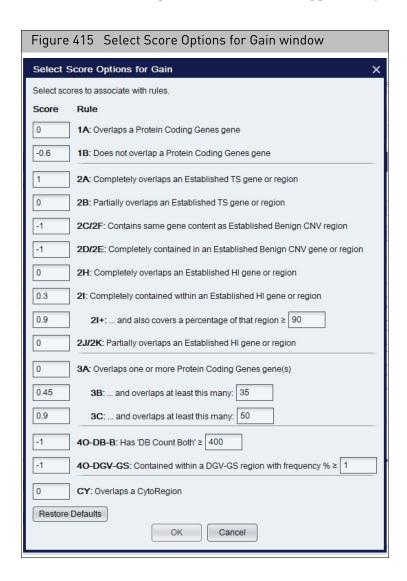
The Segment Prioritization Options window appears. (Figure 414)



2. Click on the Use Score-Based radio button.

#### View/Edit Score-Based Rules for Gain Segments

1. Click the View/Edit Score-Based Rules for Gain Segments button. The Select Score Options for Gain window appears. (Figure 415)



2. Use the text field adjacent to the Rule to enter a new numerical. Click the **Restore Defaults** button to return to the factory values. See the table below for Scorebased rule/evidence, descriptions, and default value information.

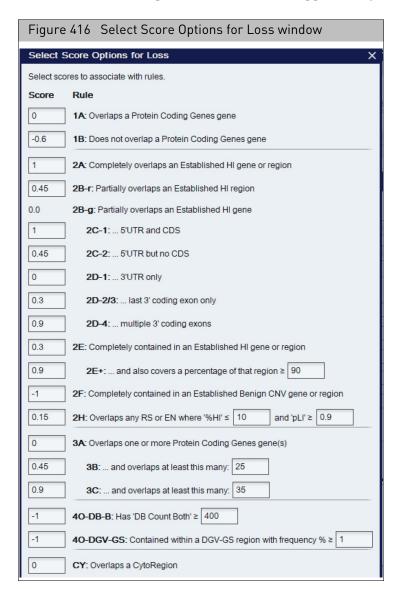
Score-based rule/evidence	Description	Default value
1A	The Gain copy number segment fully or partially overlaps at least 1 annotation in the Protein Coding Genes track.	0
1B	The Gain copy number segment does not fully or partially overlap any annotation in the Protein Coding Genes track.	-0.6

Score-based rule/evidence	Description	Default value
2A	The Gain copy number segment completely overlaps an annotation in either the Triplosensitivity or Recurrent CNV track with a TS Score = 3.	1
2В	The Gain copy number segment partially overlaps an annotation in either the Triplosensitivity or Recurrent CNV track with a TS Score = 3. Partial overlap indicates one breakpoint of the Gain segment is located within the TS_Score = 3 gene/region.	0
2C/2F	The Gain copy number segment contains the same gene content as a Triplosensitivity or Recurrent CNV annotation with a TS Score = 40. The copy number Gain segment might be larger than the gene/region, but contains the same gene content as listed in the Triplosensitivity or Recurrent CNV tracks.	-1
2D/2E	Both breakpoints of the Gain copy number segment are contained within an annotation having a TS Score = 40 in either the Triplosensitivity or Recurrent CNV.	-1
2H	The Gain copy number segment completely overlaps an annotation in either the Haploinsufficiency or Recurrent CNV track with a HI Score = 3.	0
21	Both breakpoints of the Gain copy number segment are contained within an annotation having a HI_ Score = 3 in either the Haploinsufficiency or Recurrent CNV. The copy number Gain segments is smaller than the user defined threshold (Default >=90%).	0.3
21+	Both breakpoints of the Gain copy number segment are contained within an annotation having a HI_ Score = 3 in either the Haploinsufficiency or Recurrent CNV. The copy number Gain segments is larger than the user defined threshold (Default >=90%).	0.9
2J/2K	The Gain copy number segment partially overlaps an annotation in either the Haploinsufficiency or Recurrent CNV track with a HI Score = 3. Partial overlap indicates one breakpoint of the Gain segment is located within the HI_Score = 3 gene/region.	0
3A	The Gain copy number segment (partially or completely) overlaps at least 1 Protein Coding Gene annotation. Default is 1-34 genes.	0
3B	The Gain copy number segment (partially or completely) overlaps more Protein Coding Gene annotations than in 3A. Default is 35-49.	0.45
3C	The Gain copy number segment (partially or completely) overlaps more Protein Coding Gene annotations than in 3A or 3B. Default is > =50.	0.9
40-DB-B	The Gain copy number segment overlaps/covers a defined number of segments in your ChAS database (DB Count Both column). Default is 400 segments. Configuration of DB Count Both parameters can be found in "Querying a segment from the segment table" on page 390.	-1.0
40-DGV-GS	Both breakpoints of the Gain copy number segment are contained within an annotation in the DGV-GS Gain (blue). The DGV-GS annotation must have an NR frequency greater than track defined. Default NR frequency is 1%.	-1.0
CY	The Gain copy number segment overlaps an annotation in the customer supplied CytoRegions File(s). For more information on CytoRegion files, see "Using CytoRegions" on page 268.	0

3. After your score assignments are complete, click  $\mathbf{OK}$ .

#### View/Edit Score-Based Rules for Loss Segments

1. Click the View/Edit Score-Based Rules for Loss Segments button.
The Select Score Options for Loss window appears. (Figure 415)



2. Use the text field adjacent to the Rule to enter a new value. Click the **Restore Defaults** button to return to the factory values. See the table below for Scorebased rule/evidence, descriptions, and default value information.

Score-based rule/evidence	Description	Default value
1A	The Loss copy number segment fully or partially overlaps at least 1 annotation in the Protein Coding Genes track.	0
1B	The Loss copy number segment does not fully or partially overlap any annotation in the Protein Coding Genes track.	-0.6

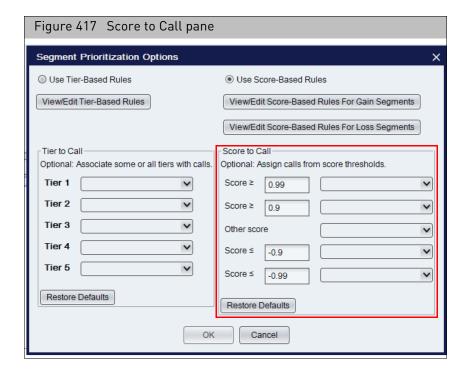
Score-based rule/evidence	Description	Default value
2A	The Loss copy number segment completely overlaps an annotation in either the Triplosensitivity or Recurrent CNV track with a TS Score = 3.	1
2B-r	The Loss copy number segment partially overlaps an annotation in the Recurrent CNV track with a HI Score = 3. Partial overlap indicates one breakpoint of the Loss segment is located within the HI Score = 3 region.	0
2B-g	The Loss copy number segment partially overlaps an annotation in the Haploinsufficiency track with a HI Score = 3. Partial overlap indicates one breakpoint of the Loss segment is located within the HI Score = 3 gene. If 2B-g is met, then move on to 2C - 2E to assess a value based on location of the partial overlap.	0 (static value, further assessment required)
2C-1	The Loss copy number segment overlaps the 5'UTR and some CDS of a gene with HI score = 3 in the Haploinsufficiency track.	0.9
	TIP: Right-click on the transcript, choose View/Edit Annotation Properties, then select the Structure tab to view the exons and CDS coordinates.	
	<b>Note:</b> All transcripts for a gene are assessed as long as the transcript is =< 90% of the length of the gene as defined in the Haploinsufficiency track and have identical gene symbols.	
2C-2	The Loss copy number segment overlaps the 5'UTR, but no CDS of a gene with HI score = 3 in the Haploinsufficiency track.	0
	TIP: Right-click on the transcript, choose View/Edit Annotation Properties, then select the Structure tab to view the exons and CDS coordinates.	
	<b>Note:</b> All transcripts for a gene are assessed as long as the transcript is =< 90% of the length of the gene as defined in the Haploinsufficiency track and have identical gene symbols.	
2D-1	The Loss copy number segment overlaps the 3'UTR only, no CDS is involved for a gene with HI score = 3 in the Haploinsufficiency track.	0
	TIP: Right-click on the transcript, choose View/Edit Annotation Properties, then select the Structure tab to view the exons and CDS coordinates.	
	<b>Note:</b> All transcripts for a gene are assessed as long as the transcript is =< 90% of the length of the gene as defined in the Haploinsufficiency track and have identical gene symbols.	
2D2/2D3	The Loss copy number segment overlaps the 3'UTR AND the last exon in the coding region for a gene with HI_score = 3 in the Haploinsufficiency track.  TIP: Right-click on the transcript, choose View/Edit Annotation Properties, then select	0.3
	the Structure tab to view the exons and CDS coordinates.  Note: All transcripts for a gene are assessed as long as the transcript is =< 90% of the length of the gene as defined in the Haploinsufficiency track and have identical gene symbols.	
2D-4	The Loss copy number segment overlaps the 3'UTR AND multiple exons in the coding region for a gene with HI_score = 3 in the Haploinsufficiency track.  TIP: Right-click on the transcript, choose View/Edit Annotation Properties, then select the Structure tab to view the exons and CDS coordinates.	0.9
	<b>Note:</b> All transcripts for a gene are assessed as long as the transcript is =< 90% of the length of the gene as defined in the Haploinsufficiency track and have identical gene symbols.	

Score-based rule/evidence	Description	Default value
2E	Both breakpoints of the Loss copy number segment are contained within an annotation having a HI_Score = 3 in either the Haploinsufficiency track or Recurrent CNV track. The copy number Loss segment is smaller than the annotation in the track by less than the user defined threshold (Default >=90%).	0.3
2E+	Both breakpoints of the Loss copy number segment are contained within an annotation having a HI_Score = 3 in either the Haploinsufficiency track or Recurrent CNV track. The copy number Loss segment is larger than the annotation in the track by less than the user defined threshold (Default >=90%).	0.9
2F	Both breakpoints of the Loss copy number segment are contained within an annotation having a HI_ Score = 40 in either the Haploinsufficiency track or Recurrent CNV track.	-1
2H	The Loss copy number segment overlaps a Protein Coding Gene or Protein Coding Ensembl Gene with predicted haploinsufficiency values meeting the defined thresholds. pLI derived from gnomAD (https://gnomad.broadinstitute.org/) and %HI derived from DECIPHER (https://decipher.sanger.ac.uk/).	0.15
3A	The Loss copy number segment (partially or completely) overlaps at least 1 Protein Coding Gene annotation. Default is 1-24 genes.	0
3B	The Loss copy number segment (partially or completely) overlaps more Protein Coding Gene annotations than in 3A. Default is 25-34.	0.45
3C	The gain copy number segment (partially or completely) overlaps more Protein Coding Gene annotations than in 3A or 3B. Default is > =35.	0.9
40-DB-B	The Loss copy number segment overlaps/covers a defined number of segments in your ChAS database (DB Count Both column). Default is 400 segments. Configuration of DB Count Both parameters can be found in "Querying a segment from the segment table" on page 390.	-0.9
40-DGV-GS	Both breakpoints of the Loss copy number segment are contained within an annotation in the DGV-GS gain (red). The DGV-GS annotation must have an NR frequency greater than track defined. Default NR frequency is 1%.	-0.9
СҮ	The Gain copy number segment overlaps an annotation in the customer supplied CytoRegions File(s). For more information on CytoRegion files, see "Using CytoRegions" on page 268.	0

3. After your score assignments are complete, click **OK**.

# Configuring the score-based option

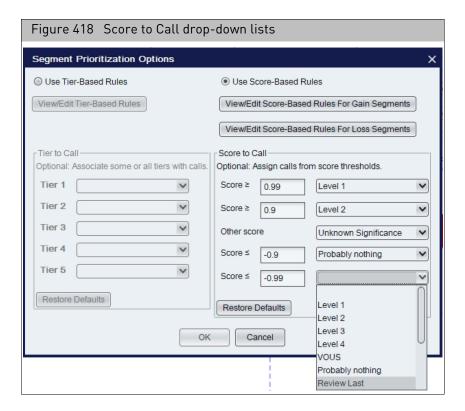
1. From the Segment Prioritization Options window, go to the Score to Call pane. (Figure 417)



- Define the Score Thresholds: In the appropriate text field, enter a Call based on the segment score as defined above. Your entered threshold values for each Call will be populated in the Segment Table's **Call from Prioritization** column.
- Select Calls: Use the drop-downs adjacent to each threshold to assign a Call that will be associated with a range of scores.

**Note:** Calls in the drop-down lists can be customized by adding to the Calls Vocabulary window in the User Configuration.

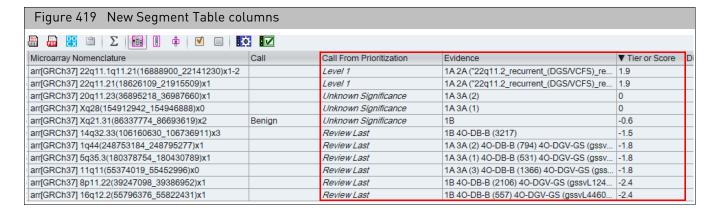
In the example below (Figure 418), a copy number segment with a Score of 1.3 would have a Call from Prioritization assignment of "Level 1". A copy number segment with a score of -0.96 would have a Call from Prioritization assignment of "Probably nothing".



2. Click **OK** to accept the Score thresholds and Calls or click Cancel to return to the ChAS Browser without saving any new assignments. Click the **Restore Defaults** button to return to the factory values.

# Viewing segment prioritization in the segments table

Three new segment prioritization columns now appear in the Segment Table. (Figure 412).



- Call From Prioritization: Displays the "Call" associated with the score threshold ranges.
- **Evidence**: Displays the rules met based on which annotations the copy number segment overlaps.

• **Tier or Score**: This will be a numeric value representing the score generated and assigned to the copy number segment based on the user-defined Score-Based rules selected.

If the Call from Prioritization assignments are correct, they can be accepted as the Calls for each segment.

#### To do this:

1. Click the button.

The Call from Prioritization assignments will be copied into blank cells in the Call column.

**Note:** Any Calls manually assigned will remain in the Call column and not be overwritten. Segments hidden by the filters will not have calls copied from the Calls from Prioritization column.

A confirmation message appears (Figure 420) summarizing the Call from Prioritization assignments are to be copied into the Call column.



2. Click Yes to acknowledge the message.



# Interacting with the ChAS database

#### This chapter includes:

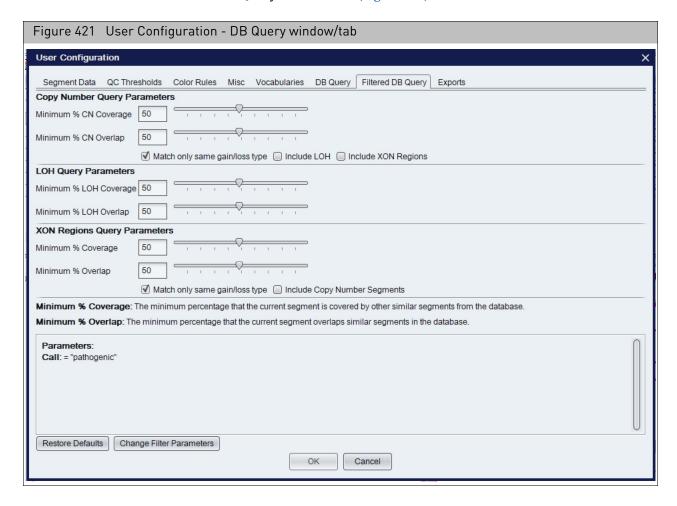
- "Setting up a ChAS DB query" on page 388
- "Querying a segment from the segment table" on page 390
- "Additional segment intersection information" on page 395
- "Filtering DB count columns" on page 399
- "Querying a segment from the detail view" on page 401
- "Changing or refining the DB query criteria" on page 402
- "Publishing data to the database" on page 404
- "Manual or automatic connection mode" on page 406
- "Querying samples in the ChAS database" on page 407
- "Removing a sample from the query window" on page 409
- "Deleting sample(s) from the ChAS database" on page 409
- "Querying segments to the ChAS database" on page 411

A segment can be queried against the ChAS Database for intersecting segments from previously published samples. Using both the Overlap Threshold and Coverage Threshold can focus the query results to segments that are of approximately the same size as the segment in the current sample.

**Note:** ReproSeq Aneuploidy data can not be published into the ChAS DB.

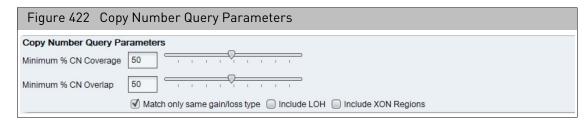
### Setting up a ChAS DB query

- 1. From the ChAS Browser, click **Preferences** → **Edit User Configuration**.
- 2. Click the **DB Query** window tab. (Figure 421)



Setting up query parameters for a copy number search **Note:** When querying on a copy number segment in the Browser, the values set in the Copy Number Query Parameters section are used.

1. Enter minimum percentage values for both Overlap and Coverage using the text boxes or click and drag the appropriate slider bar. (Figure 422) The default values are set to 50%.

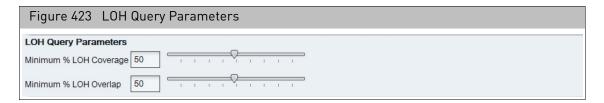


- 2. Check the **Match only same gain/loss type** check box if you want to query the database for only similar copy number types (gains only or losses only). Uncheck this check box if you want to query all copy number segment types.
- 3. Check the **Include Exon Regions** check box if you want to include Exon Region Segments in your query.
- 4. Check the **Include LOH** box to include LOH segments in your query.
- 5. Click **OK** to save your changes or click **Restore Defaults** to return the parameter settings back to their default settings.

Setting up query parameters for an LOH segment search

**Note:** When querying on an LOH segment in the Browser, the values set in the LOH Query Parameters section are used.

1. Enter minimum percentage values for both Overlap and Coverage using the text boxes or click and drag the appropriate slider bar. (Figure 423) The default values are set to 50%.

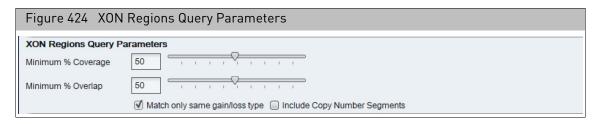


2. Click **OK** to save your changes or click **Restore Defaults** to return the parameter settings back to their default settings.

Setting up query parameters for an XON region segment search

**Note:** When querying on an XON Region segment in the Browser, the values set in the Exon Region Query Parameters section are used.

1. Enter minimum percentage values for both Overlap and Coverage using the text boxes or click and drag the appropriate slider bar. (Figure 424) The default values are set to 50%.



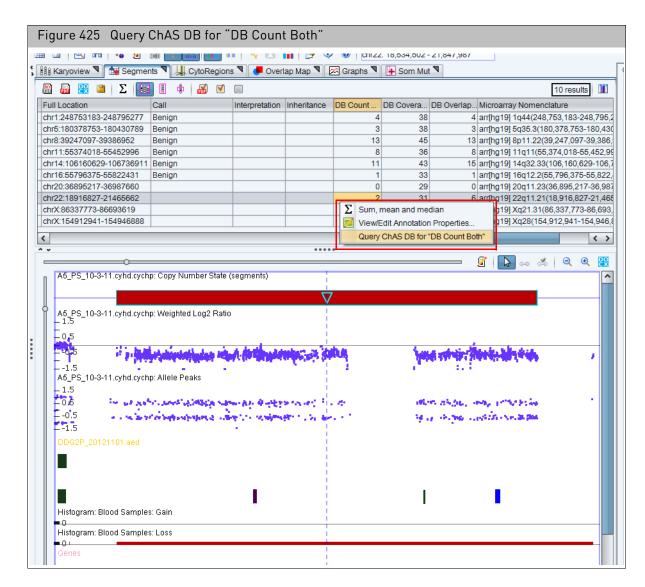
- 2. Check the **Match only same gain/loss type** check box if you want to query the database for only similar XON Region segment types (gains only or losses only). Uncheck this check box if you want to query all XON Region segment types.
- 3. Check the **Include Copy Number Segments** check box to include Copy Number Segments in your query.
- 4. Click **OK** to save your changes or click **Restore Defaults** to return the parameter settings back to their default settings.

### Querying a segment from the segment table

To view what segments in the database intersect with the currently loaded segment, you must first make the **DB Count Both** column visible. The DB Count Both column displays the number of segments in the database that meet the criteria set in the DB Query tab.

1. Right-click in the DB Count Both cell for the segment(s) you want to view, then click **Query ChAS DB for "DB Count Both"**. (Figure 425)

**Note:** Segments matching ONLY the Coverage OR Overlap thresholds can also be returned in the Segment Intersections window by right clicking in either the DB Coverage Count or DB Overlap Count columns respectively. You can also right click on the segment in the Detail View and choose Query ChAS DB to return segments meeting either the Coverage OR the Overlap thresholds. See "Filtering DB count columns" on page 399. If you want to add column(s) to your Segment table, see "Selecting columns to display or hide" on page 329.



# Segment intersections

The **Segment Intersections** view appears with the results from the query. (Figure 426)

The Segment Intersections view shows samples in the database that contain segments that meet the criteria set in the DB Query.

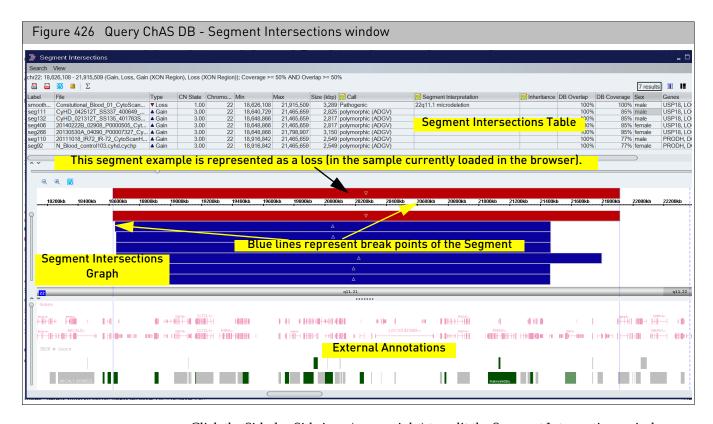
The middle portion of this view contains table information about the samples in the database including any Call, Interpretation or Inheritance information assigned to the segments for the samples shown in the example above. To display or hide columns within this table, click (upper right corner).

The lower portion of the view provides the same external annotations available in the Detail View. To display an annotation track, go to the ChAS Browser's Files Menu and check the box. The annotation track will then be displayed in both the Detail View and the Segment Intersections View. (Figure 426)

You can return segments from the database based on either DB Overlap or DB Coverage. These segments meet either one of the overlap or coverage threshold criteria.

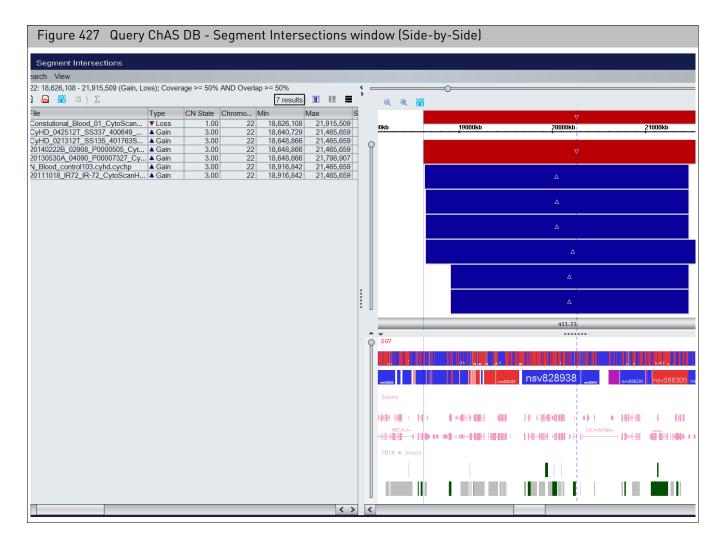
**Note:** Segment Intersection search results are limited to 1000 intersecting segments. When querying on a copy number segment in the Browser, the Copy Number Query Parameter Thresholds are used for all segment types. When querying on an LOH segment in the Browser, the LOH Query Parameter Thresholds are used for all segment types.

Example: You have a gain segment in the ChAS Browser and run a query to retrieve intersecting segments. Both Gain and LOH segments counts appear in DB Count Both, the LOH segment being returned are using the Copy Number Query Parameter Thresholds since the original segment is a GAIN.



Click the Side-by-Side icon (upper right) to split the Segment Intersections window, as shown in Figure 427 on page 393.

**Note:** Columns with a pad and pencil icon represent a segment field that can be edited. All edits are stored directly to the database.



The Segment Intersection View table has the following columns:

**Note:** Columns listed with an adjacent [3] icon denote the column is user-editable.

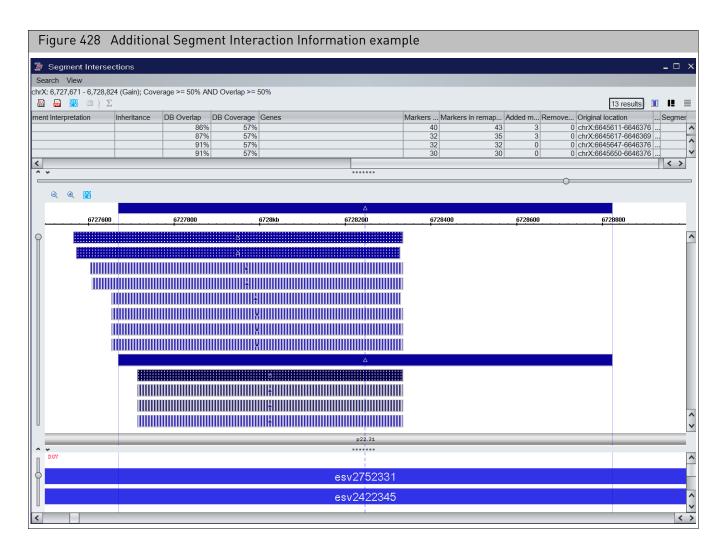
Column	Description
<b>[</b> Call	User-editable field populated by a user-configurable drop list of Calls.
Inheritance	User-editable field populated by user-configurable drop-down list of Inheritance.
oncomine Reporter	Displays the Oncomine Reporter terminology assigned to the segment.
Segment Interpretation	Displays Segment Interpretation assigned to this sample.
% of Overlap Item covered by Segment	The percentage of the Overlap Map Item covered by the segment.
Call from Prioritization (Stored)	The Call term assigned based on Tier or Score Classification at the time the sample was published to ChAS DB.

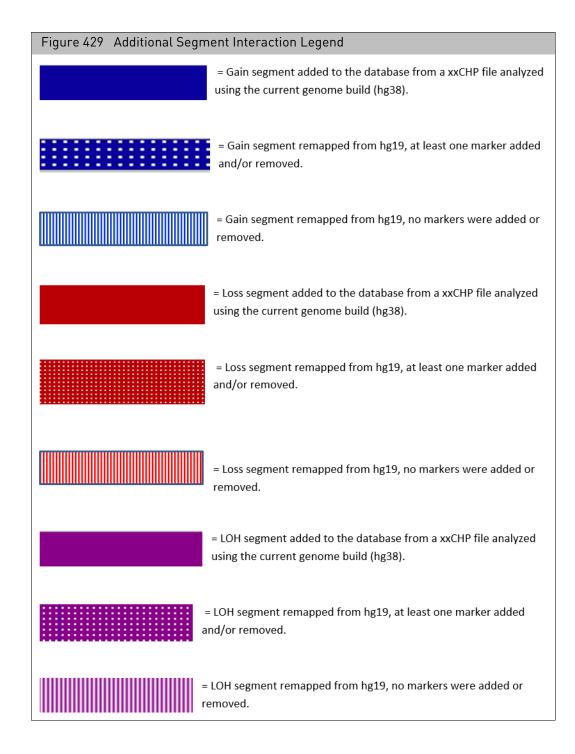
Column	Description
Chromosome	Chromosome on which the item is located.
CN State	Copy Number State.
CytoRegions	Names of the CytoRegions with which the segment shares coordinates.
DB Coverage	Number of segments in the database meeting the minimum Percent Coverage Count.
DB Overlap	Number of segments in the database meeting the minimum Percent Overlap Count.
DGV	List of DGV variations that share coordinates with the segment.
Evidence (Stored)	Provides information on which annotations the segment overlapped at the time the sample was published to ChAS DB.
Genes	List of RefSeq genes from the Genes track that share coordinates with the segment. Identically named gene isoforms are NOT repeated.
Interpretation	User-editable field for free-text interpretation on the segment.
Label	Identifier for the item.
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
Min	Zero-based index position of the first base pair in the sequence.
OMIM Genes	List of OMIM Genes that share coordinates with the segment.
Overlap Map	Item(s) in the Overlap Map which overlap the segment.
Overlap Map Items	The percentage of the segment that is overlapped by the Overlap Map Item.
Phenotype	Displays Phenotype annotation assigned to this sample.
Published	Displays the date and time of your query.
Publisher	Displays Publisher's name.
Sample DB ID	A xxCHP file ID automatically assigned when the xxCHP file is published to the database.
Sample Type	Displays the Sample Type assigned to this xxCHP file.
Sample UUID	Unique identifier for the CHP file.
Segment DB ID	An ID automatically assigned to each segment when the xxCHP file is published to the database
Segment ID	The unique identifier for the copy number segment.
Segmental Duplications	List of Segmental Duplications that share coordinates with the segment.
Sex	Displays Male or Female.
Size (kpb)	Size of the item.

Column	Description
Tier or Score (Stored)	The assigned Tier or Score value based on the Segment Prioritization method selected at the time the sample was published to ChAS DB. When using Tier based, the column will display the assigned Tier. When using Score based, the column will display the score value based on the annotations the segment overlaps.
Туре	Type of segment (Gain, Loss, GainMosaic, LossMosaic LOH) or annotation.

### Additional segment intersection information

Additional Segment Intersection information becomes available after querying a ChAS DB that has been remapped from a previous genome build that includes additional columns that are populated in the Segment Intersection Table. These remapped segments, are also represented in different graphical patterns, as shown in Figure 428 and detailed in Figure 429.





The additional Segment Intersection table columns and their definitions, are as follows:

Column	Description			
Original location	Chr:start-stop genome location of the original genome build for the segment.			
Markers in original segment	Total number of markers in the original segment.			
Added markers	Number of markers to the segment added by remapping.			
Removed markers	Number of original markers removed from the segment from remapping.			
Segment length difference	Size difference in the segment (Original - Remapped).			
Original Genome Build	Genome Build from the original analysis prior to remapping.			
Markers in remapped segment	Total number of markers in the remapped segment.			

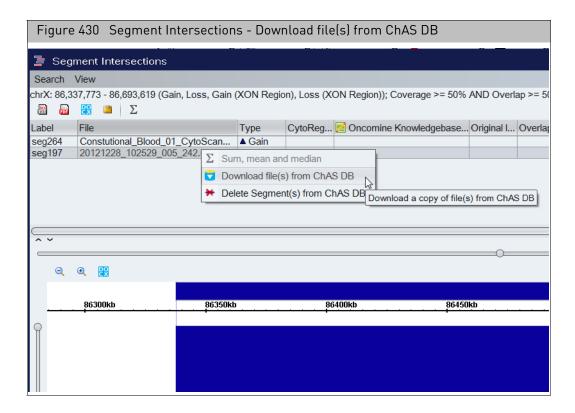
## Downloading segments from a sample file in ChAS DB

From the Segment Intersection window, all the segments for a selected sample can be downloaded and viewed in the Segments Table and Detail View. Only the segment data and annotations from the sample are displayed.

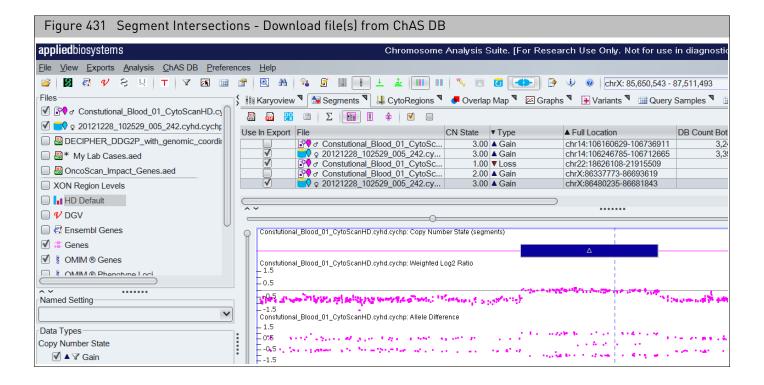
**Note:** Files downloaded from ChAS DB can not be opened in the MSV.

To Download and View Segments from a sample(s) stored in ChAS DB:

- 1. From the Segment Intersections window, right-click on the sample(s) in either the table or the graphical view.
- 2. Select Download file(s) from ChAS DB. (Figure 430)



**Note:** As shown in Figure 431, sample files downloaded from ChAS DB are listed in the Files Tree with a database symbol. Segments from samples files downloaded from ChAS DB are listed in the Segments Table with a database symbol in the File column.



**Note:** Downloaded segments can be deleted when in Edit Mode, but no other segment modifications are enabled as the underlying data is unavailable.

### Updating downloaded segment annotations

Segment annotations in the ChAS DB can be updated using the following methods:

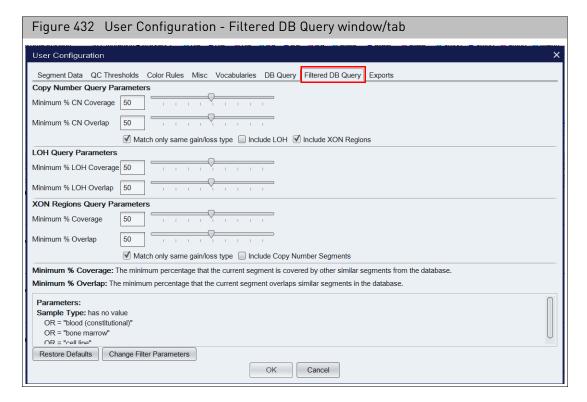
- Right-clicking on the segment and selecting View/Edit Annotation Properties → Curation Tab.
- Clicking on a column/field pad and pencil icon to edit the segment annotation(s).
- Right-clicking on the Filename in the Files tree, select View/Edit Properties → Sample Properties.

## Filtering DB count columns

Filtered DB count columns are available in the Segment Table and reflect the number of segments in the database matching the filtered criteria.

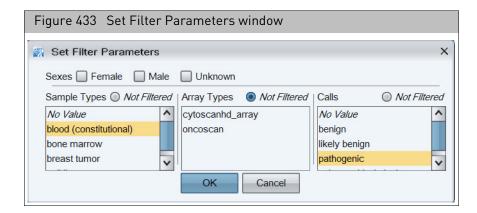
**Note:** The DB count columns reflect the number of segments in the database meeting the Minimum Percent Overlap/Coverage criteria only. The Filtered DB count columns allow you to display segments that not only meet the Minimum Percent Overlap/Coverage criteria, but also additional filters such as Call or Gender.

1. Click **Edit User Configuration**, then click on the **Filtered DB Query** tab. (Figure 432)



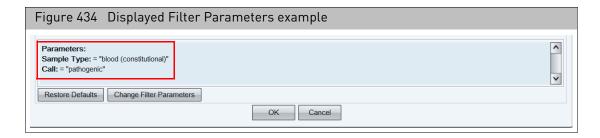
- 2. Refer to "Setting up a ChAS DB query" on page 388 for how to set the Percent Minimum Overlap/Coverage Thresholds.
- 3. Return to the Filtered DB Query window tab.
- 4. The Filtered DB Query window tab displays the current filter setting. To change it, click **Change Filter Parameters**.

A Set Filter Parameters window appears. (Figure 433)



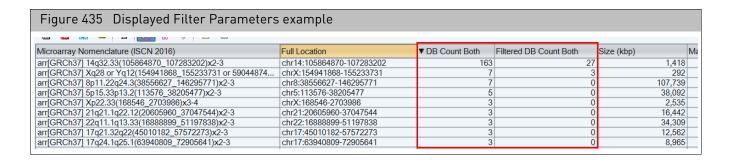
- 5. Use the windows check boxes, radio buttons, and selections to change your filter parameters.
- 6. Click OK.

Your new query parameters are saved and displayed at the bottom of the Filtered DB Query window tab, as shown in Figure 434.



To reset the Filtered DB Query Parameters back to default settings, click **Restore Defaults**. For more information see, "Filtering DB count columns" on page 399.

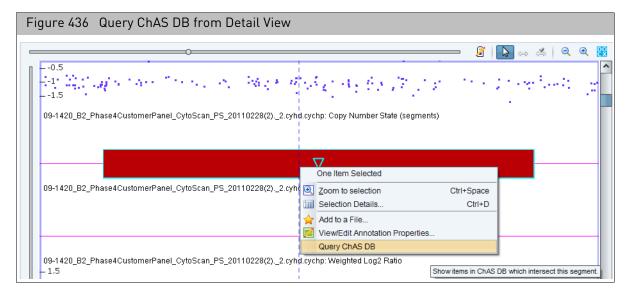
Figure 435 shows the **DB Count Both** and **Filtered DB Count Both** columns. It illustrates that DB Count Both queries the database for all segments matching the Minimum Percent Overlap/Coverage and gain/loss/LOH parameters. The Filter DB column reflects the additional Filter Criteria.



## Querying a segment from the detail view

To view what segments in the database intersect with the currently loaded segment:

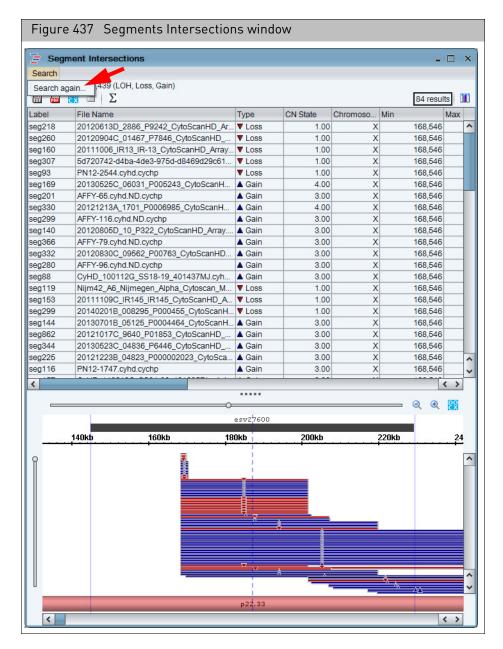
- 1. Right-click on a Segment of interest within the Detail View. (Figure 436)
- 2. Click Query ChAS DB.



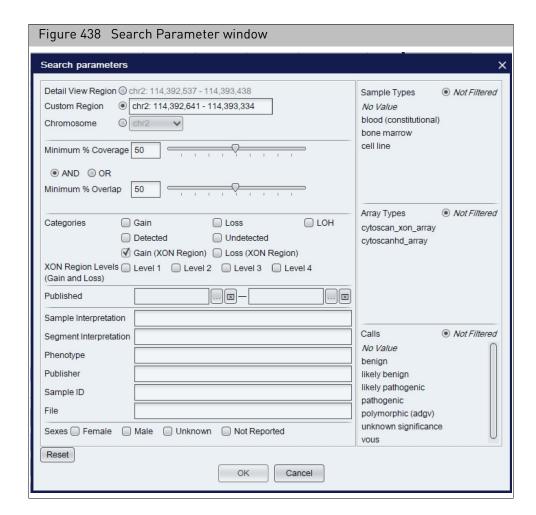
The results returned when querying a segment in the detail view will contain segments that meet the DB Coverage filter or the DB Overlap filter set up previously. (Figure 421 on page 388)

## Changing or refining the DB query criteria

1. From the upper left corner, click **Search** → **Search Again...** (Figure 437)



The Search parameters window appears. (Figure 438)



2. Use the provided radio buttons, check boxes, and text fields to customize your search, then click **OK**.

**Note:** Altering these parameters only affects the current segment query. The following fields, **Sample Types**, **Array Types** and **Calls** are populated based on what has been published by the user into the ChAS database If the ChAS Browser is unable to contact the database, these fields are populated based on the library file and vocabularies entries.

### Publishing data to the database

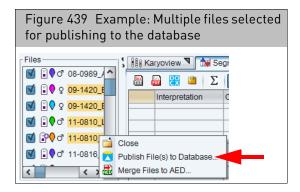
Once sample and its segments have been analyzed, curated and annotated, it can be added to a database by a process called publishing.

**IMPORTANT!** You MUST have Manager or Admin Role permissions to publish data to the database. For details, see "Administration" on page 454.

Publishing data or multiple data to the database

#### Method 1

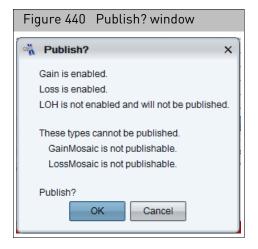
1. In the File tree, right-click on a file name, then click **Publish File(s) to Database...** (Figure 439)



#### Method 2

1. Click to highlight the sample name(s), then click the tool bar's Publish to Database icon .

A summary of uploaded segments/Publish? appears. (Figure 440)



2. Acknowledge the message, then click **OK**.

**Note:** Only segments in the copy number, XON region and LOH segment tracks can be uploaded. Segments uploaded are reflective of the current filters and settings applied.

**Note:** Only segments listed as enabled and filtered will be published. If a segment category is listed as not enabled, click Cancel and check the check box in the Data Types File tree and then right click on the sample filename(s) to start the publishing again.

**Note:** Publishing time is dependent on the number of segments in the sample.

**Note:** Segments in the Mosaic track are not uploaded to the database. Mosaic segment must be promoted to the copy number state track in order to be published to the database. See "Promoting mosaic segments" on page 244.

**Note:** If a xxCHP file has been previously Published to the database, you will receive a warning indicating this sample already exists in the database. You can choose to overwrite the existing information or cancel to keep the existing information.

**Note:** The Segment Table Columns DB Count and Filtered DB Count are automatically updated.

**Note:** If ChAS is set to manual mode, the histogram need to be manually updated to include recently published samples. To update the histograms, click **ChAS DB** → **Refresh ChAS DB** data.

## Publishing to database

#### Important rules and restrictions

- A sample cannot be published if it is in Edit Mode. See "Using edit mode" on page 226.
- Only samples that are highlighted (not checked) are published to the database.
- Samples published using hg18 cannot be published to the database.
- Multiple filenames can be highlighted and published at the same time.
- Segments from xxCHP files analyzed using hg19 analysis files cannot be loaded into a ChAS Database genome version hg38. The opposite is also true.
- Mosaic Segments can not be published.
  - To promote mosaicism in the database, the Mosaic Segment should be assigned a copy number value and promoted to Copy Number segments using the edit function. For more details, see "Promoting mosaic segments" on page 244.
- You must be logged in as an administrator or manager (with manager role permissions) to publish data.
- Previously published samples can be edited and published again, however publishing a second time overwrites the original database entry.

### Manual or automatic connection mode

By default, ChAS starts in automatic connection mode.

While in automatic connection mode, the DB count columns and histograms are refreshed/updated whenever a file is published.

If you do not want to wait for the data to be refreshed/updated each time a file is published, choose manual connection mode. In manual mode, the DB count columns and histograms are only be updated when you click  $ChASDB \rightarrow RefreshChASDB$  data or click the representation.

There are two ways to switch from automatic to manual connection mode.

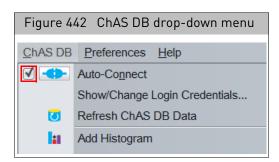
• At start up, select a user (as you normally would), click on the **Manual connection** check box, then click **OK**. (Figure 441)



• During a ChAS session, go to the upper bar of icons and click on automatic connection or click on for manual connection.

Or

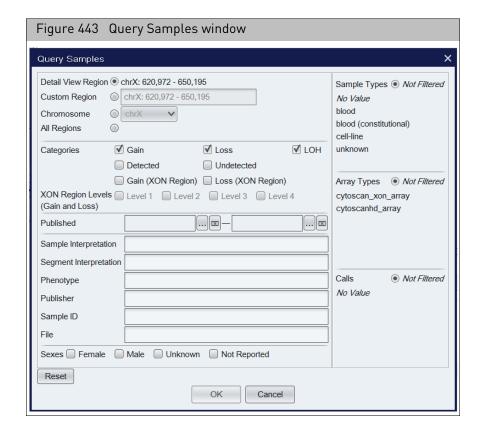
Click **ChAS DB**  $\rightarrow$  **Auto-Connect**, then click on the check box to toggle between connection modes, as shown in Figure 442.



## Querying samples in the ChAS database

 Click ChAS DB → Query Samples or click the Query Samples tab, then click the Query Samples button.

The **Query Samples** window appears. (Figure 443)



2. Use the provided radio buttons, check boxes, and text fields to customize your query, then click **OK**.

The Query Samples table populates with your filtered search results. (Figure 444)

The contents for columns in which the headers have an Edit Icon can be modified. Changes will apply directly to the ChAS database.

Columns with a chip/pencil [2] icon represent sample properties that can be edited.

## Editing column contents

- Double click on a cell to be edited.
   An Edit Value window appears. (Figure 444)
- 2. Enter or change the value in the window.
- 3. Click OK.

Your changes are now saved to the ChAS Database.

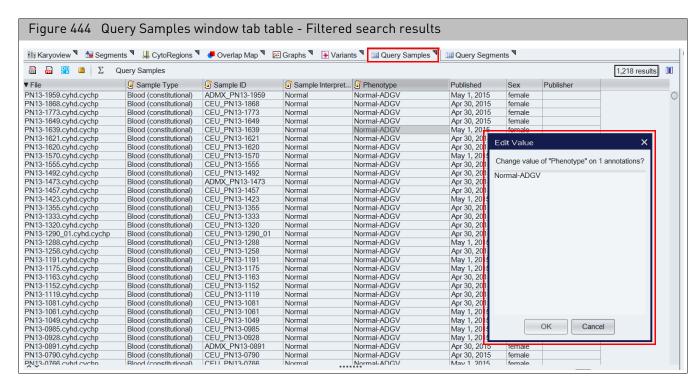
# Editing multiple cells with the same value

- 1. Highlight the cells you want to edit.
- 2. Right-click, then select **Edit Property Values**.

An Edit Value window appears. (Figure 444)

- 3. Enter a new value or edit the existing displayed value.
- 4. Click OK.

Your changes (for the multiple cells you selected) are now saved to the ChAS Database.



For instructions on how to use the table's features, see "Common table operations" on page 327.

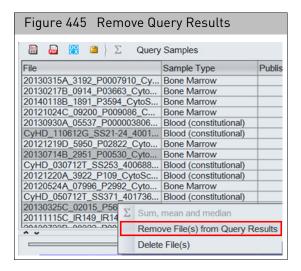
**Note:** The following fields, **Sample Types**, **Array Types** and **Calls** are populated based on what has been published by the user into the ChAS database. If the ChAS Browser is unable to contact the database, these fields are populated based on the library file and vocabularies entries. Queries are not automatically refreshed when publishing to or deleting from the ChAS DB. Queries must be re-run to reflect changes to the database made after the initial query.

## Removing a sample from the query window

You can remove a sample from the query window, however this action does NOT remove the sample from the database.

- 1. Highlight the sample(s) you would like to remove from the results display.
- 2. Right-click, then click **Remove Query Results**. (Figure 445)

  A warning message appears asking you to confirm the removal of the file(s) from the results window.
- Click **OK** to remove the files from the results. Click **Cancel** to return to the main screen.



## Deleting sample(s) from the ChAS database



**WARNING!** You must have manager or admin permissions for the ChAS database to delete samples.

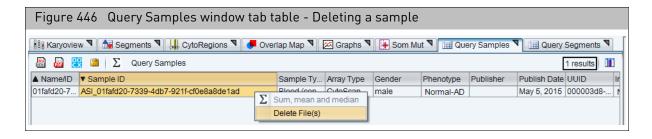
Sample(s) deleted from the ChAS Database are permanently deleted and cannot be retrieved. There is no undo delete feature.

## Deleting a single sample

To remove a single sample in a database, use the Query Samples window to locate the file to be deleted.

1. In the Query Samples window (Figure 443 on page 407), enter the file's **Filename** or **Sample ID**, then click **OK**.

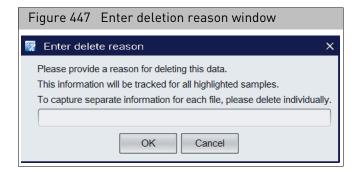
The sample appears in the table. (Figure 446)



2. Right-click on the sample, then click Delete.

When a sample is deleted from a database, the reason for the deletion is required. Enter the reason in the **Enter delete reason** window. (Figure 447)

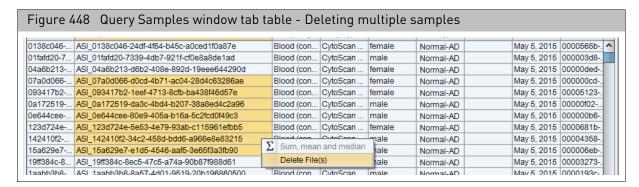
**Note:** This reasoning you enter is captured in the ChAS DB and it can be exported. For details, see "Exporting a deletion log" on page 435.



3. Click **OK** to delete. Click **Cancel** to return to the query window. The sample is removed.

## Deleting multiple samples

 Multiple samples can be highlighted to delete. They can be selected using the following keyboard and mouse combinations: Ctrl click, Shift click or Ctrl a.. (Figure 448)

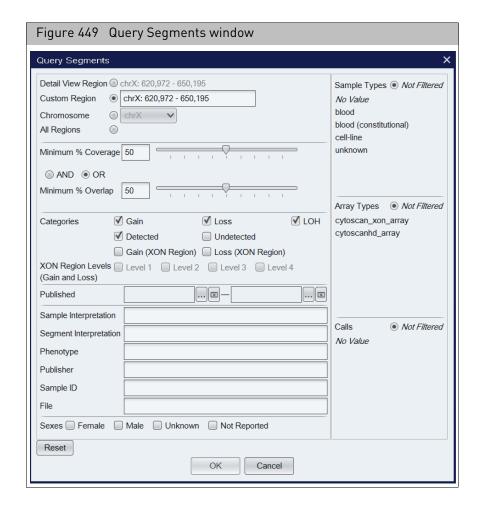


2. Right-click on the highlighted area, then click **Delete**. The samples are removed.

## Querying segments to the ChAS database

1. Click ChAS DB → Query Segments or click the Query Segments tab, then click the Query Segments button.

The **Query Segments** window appears. (Figure 449)



2. Use the provided radio buttons, check boxes, and text fields to customize your query, then click **OK**.

The Query Segments table populates with your filtered search results. (Figure 450)

The contents for columns in which the headers have an Edit Icon can be modified. Changes will apply directly to the ChAS database. Editable columns are: Call, Segment Interpretation, Inheritance, Oncomine Reporter.

## Editing column contents

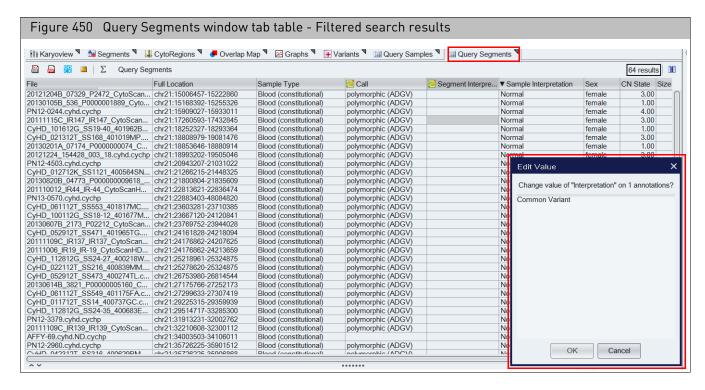
- Double click on a cell to be edited.
   An Edit Value window appears. (Figure 450)
- 2. Enter or change the value in the window.
- 3. Click OK.

Your changes are now saved to the ChAS Database.

# Editing multiple cells with the same value

- 1. Highlight the cells you want to edit.
- 2. Right-click, then select **Edit Property Values**. An Edit Value window appears. (Figure 450)
- 3. Enter a new value or edit the existing displayed value.
- Click OK.

Your changes (for the multiple cells selected) are now saved to the ChAS Database.



For instructions on how to use the table's features, see "Common table operations" on page 327.

**Note:** To delete sample files from the Query Segments tab, follow the same instructions outlined in "Deleting sample(s) from the ChAS database" on page 409. This procedure deletes the entire sample file and all file information associated with it.



## **Exporting results**

#### **IMPORTANT!** Edit Mode must be OFF before exporting from ChAS.

Chromosome Analysis Suite includes the following tools for reporting results:

- "Exporting graphic views": Export the Karyoview, Selected Chromosome View, and Detail View as a DOCX, PDF or PNG file.
- "Creating signature and background profiles": Export to a DOCX file.
- "Exporting table data": Export table data as a DOCX, PDF, TXT file, or copy selected data onto your clipboard.
- "Combining PDFs into a single PDF"
- "Exporting with ClinVar": Using a ClinVar export template.
- "Exporting VCF files" on page 325: Export copy number and variant data as VCF.

**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" folder for the file ARIALUNI.TTF.

### **Exporting graphic views**

You can export the Karyoview, Selected Chromosome View, and Detail View in the following formats:

- PDF "Exporting as a PDF"
- MS Word (DOCX) "Exporting as Word (DOCX) format" on page 418
- PNG graphic file "Exporting as PNG" on page 420

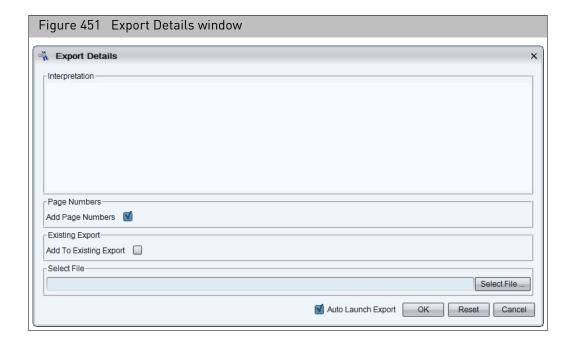
#### Exporting as a PDF

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.

- 1. From the Exports menu, select the PDF option you want 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

An Export Details window appears. (Figure 451)



- 2. Optional: If you have added a Sample Interpretation in the View/Edit Sample Properties window, the information will be populated into the Interpretation dialog box. You can also type free text into the Interpretation box.
- 3. Optional: Select the option for adding page numbers.
- 4. Optional: Select the option for adding to an existing report, if desired. See "Combining PDFs into a single PDF" on page 428.
- 5. Click **Select File**.

The **Save** window opens.

- Select a folder location for the PDF using the navigation tools.
   This folder location is automatically selected when exporting other PDFs during a session.
- 7. Enter a name for your PDF file.

If you are adding the graphic to an existing PDF, select the PDF file.

8. Click Save.

You are returned to the **Export Details** window.

- 9. Optional: Leave the Auto Launch Export check box checked, if you want your newly saved PDF to open automatically after clicking **OK**.
- 10. Click OK.

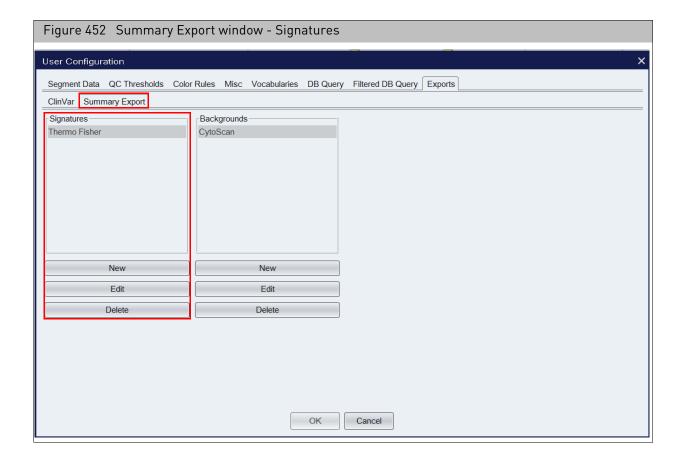
A PDF is created with your selected export details saved.

### Creating signature and background profiles

#### Signature profiles

Signature Profiles, including a logo, address, reviewer name and credentials, can be added to a DOCX export. Use saved signature profiles for quick recalls with any DOCX export

- 1. Click on the Preference Menu, then select **Edit User Configuration**.
- 2. Click the **Exports** Tab.
- 3. Click the **Summary Exports** tab, then click the **Summary Export** tab. (Figure 452)



#### Creating a new signature

- 1. Click the **New** button.
- 2. Complete the text fields.
- 3. Optional: Click on the **Upload Logo** button to add your organization's logo.
- 4. Click **OK** to save.
  - Your saved signature name will be the same as your organization name.
- 5. Optional: Repeat steps 1-4 to create additional signatures.

#### Editing a saved Signature

1. Highlight the Signature name you would like to modify.

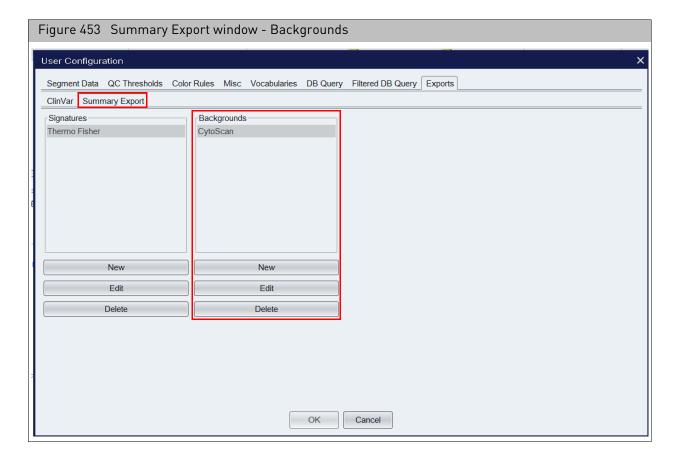
- 2. Click the Edit button.
- 3. Modify the appropriate fields.
- 4. Click **OK** to save your updated signature.

#### Deleting a signature

- 1. Highlight the Signature name you would like to delete.
- 2. Click the **Delete** button.
- 3. Click **OK** to permanently remove the signature.

## Background profiles

Background profiles (Figure 453) provide saved text(s) that can be added to each DOCX export. For example, a noteworthy background about the assay profiles you want to save.



#### Creating a new background

- 1. Click on the **New** button.
- 2. Name the Background using the Title field.
- 3. Enter the text for the background information that you want to appear on the export.
- 4. Click **OK** to save.
- 5. Optional: Repeat steps 1-4 to create additional Backgrounds.

#### Editing a saved background

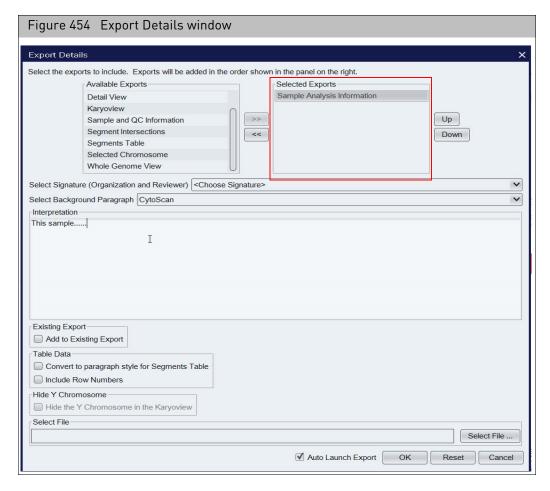
- 1. Highlight the Background name you want to modify.
- 2. Click the Edit button.
- 3. Modify the text.
- 4. Click **OK** to save your updated background.

#### Deleting a background

- 1. Highlight the Background name you would like to delete.
- 2. Click the **Delete** button.
- 3. Click **OK** to permanently remove the Background.

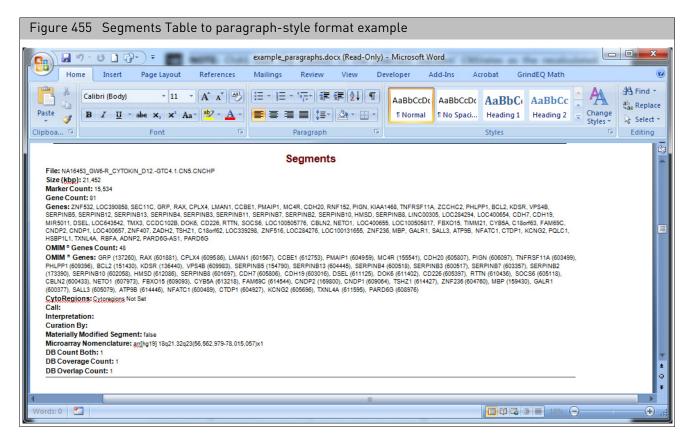
## Exporting as Word (DOCX) format

1. From the Exports menu, select **Export - Word (docx) Format**. The **Export Details** window opens. (Figure 454)



**Note:** In order to export a graph (Karyoview, Whole Genome View, Detail View, Selected Chromosome View), the graph must be visible in the ChAS Browser before export.

- Single-click, Ctrl-click, or Shift-click on the Available Exports (left pane), then click
  the right double arrow button (Figure 454) to add them to the Selected Exports (right
  pane). Use the Up and Down buttons to define each report's order within the master
  report.
- 3. Optional: Select a saved Signature and/or Background from the drop-down lists to add them to your DOCX export.
  - If you previously entered Sample Level Interpretations, they will appear in the Interpretation field.
  - **Note:** Before adding an entry in the Interpretation field, the Selected Exports pane must first display **Sample Analysis Information**, as shown in Figure 454.
- 4. Optional: Click the **Add to Existing Export** check box to add this new report to an existing one. After checking this check box, click on the **Select File** button to navigate to and select an existing DOCX file.
- 5. Optional: Click the **Convert to paragraph style for Segments Table** check box if you want your Segments Table data translated into a paragraph-style format, as show in Figure 455.



- 6. Optional: Click the **Include Row Number** check box to add row numbers to the Segment Table.
- 7. Optional: Click the **Hide Y Chromosome** check box to export the Karyoview without the Y chromosome ideogram for female samples.
- 8. Click the **Select File** button to navigate to a saved report location, enter a filename, then click **Save**.

The name of the file defaults to the name of the filename.

- 9. By default, the report automatically opens in MS Word after it is generated. Uncheck the **Auto Launch Export** check box to disable this feature.
- 10. Click OK to generate the DOCX report. Click the Reset button to return the Report Details window back to its factory defaults.

#### **Exporting as PNG**

You can also create a PNG screen shot of the entire software screen.

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

- From the Exports menu, select Export application window PNG.
   A Save As window opens.
- Select a folder location for the file using the navigation tools.This folder location is automatically selected when exporting other screen shots 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:

- "Exporting table data into a PDF" on page 420
- "Exporting views and tables as a DOCX file" on page 423
- "Exporting tables as TXT file" on page 423
- "Exporting a segments table with modified segments to a TXT file" on page 425

## Exporting table data into a PDF

For information on how to choose and export preset column content (from previously saved table states). See "Saved table states" on page 332.

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

**Note:** You can export data from the Segment Table by selecting Export Segment Table PDF from the Exports menu, but you cannot export graph table data in a PDF format.

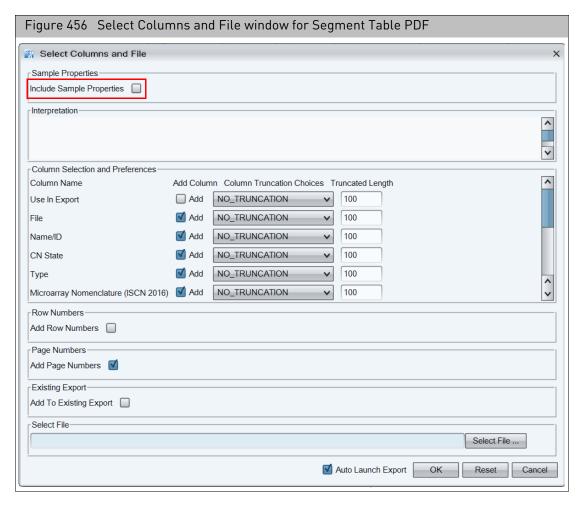
In order to track which of your segments have been modified (merged, created de novo, deleted, had their start or end coordinates edited, or had their type or state changed), you **MUST** perform the following **BEFORE** exporting the Segments Table to PDF. The only PDF format capable of tracking which segments have been modified is the Segments Table PDF, all other PDFs show NO visual or textual distinction between modified and non-modified segments.

BEFORE exporting table data into a PDF file:

- Click (top right of Segments table).
   The Select Columns window appears.
- 2. Scroll down, locate, then click to check the **Modified Segment** check box.
- Click OK.
   The Modified Segment column is now added to your Segments table.

#### Exporting table data into a PDF file

1. At the Segments table upper tool bar, click the PDF button The Select Columns and Files window appears. (Figure 456)



If you have added a Sample Interpretation in the View/Edit Sample Properties window, the information will be populated into the Interpretation dialog box. You can also type free text into the Interpretation box. You must check the **Include Sample Properties** check box in order to enable the Interpretation field

- Enter the appropriate text in the Interpretation field.
   To add the other Sample Properties, check the Include Sample Properties check box. If these fields have been populated, they will be exported in the PDF.
- 3. Select the columns to be displayed and the columns truncation rules. The Column Selection and Preferences window (Figure 457), include:

- Column Name: Header of the column in the table.
- Add Column: Click the check box to display the column in the PDF report.
- Column Truncation Options:

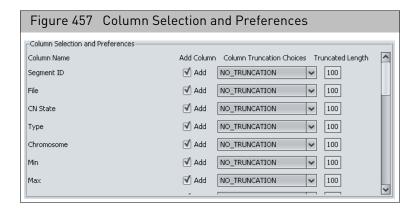
**No\_Truncation** - Field is exported as is, using wrap-around if necessary.

**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.

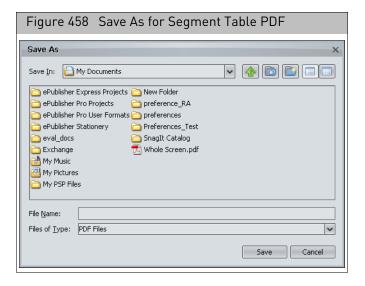
**Truncate\_End** - Truncates content at the end of the field, leaving as many characters as specified in Truncated Length box.

- Truncated Length: Number of characters displayed after truncating the data.



- 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 PDF" on page 428.
- 6. Click the **Select File** button.

The Save As window opens. (Figure 458)



- 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.
- Click Save in the Save As window.
- 10. Click **OK** in the **Select Columns and File** window.

A PDF file is created with the selected data type saved.

The PDF report displays:

- Table type
- Information on chromosome and genome region
- Interpretation
- Data files
- Genome or CytoRegion Segment Filters used
- Settings for Data Processing
- Microarray Nomenclature
- Details of the table data

#### Exporting tables tips

Follow the tips below to improve the export of table data in a PDF file:

- Use truncation.
- Select only columns you need or select columns from a saved table state. See "Saved table states" on page 332.
- 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.

#### Exporting views and tables as a DOCX file

You can save the currently viewed table in a new or existing DOCX report.

1. While in a desired View or Table, click the DOCX button A Save window appears.



Do one of the following:

- New Reports: Enter a filename to create a new DOCX file, then click Save.
- Adding to an existing report: Click to select an existing DOCX file. Click Save. The message window, You will be adding contents to an existing file. Continue? appears. Click Yes to append the currently displayed table contents onto an existing file.

#### IMPORTANT! Before adding a file, it must be closed.

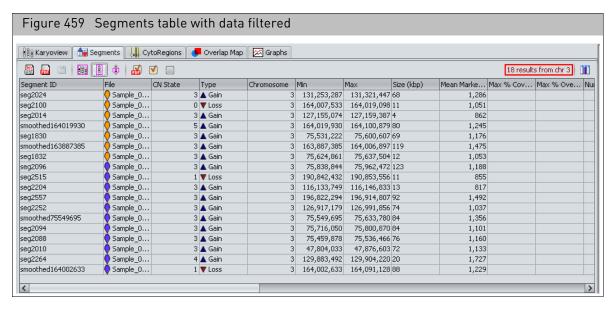
#### Exporting tables as TXT file

The TXT file format enables you to transfer data to other software for analysis.

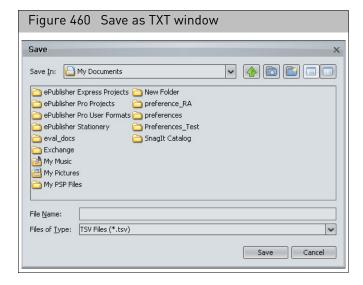
IMPORTANT! The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.

#### Exporting table information as a text file

1. Perform pre-filtering on the data in the table. (Figure 459)



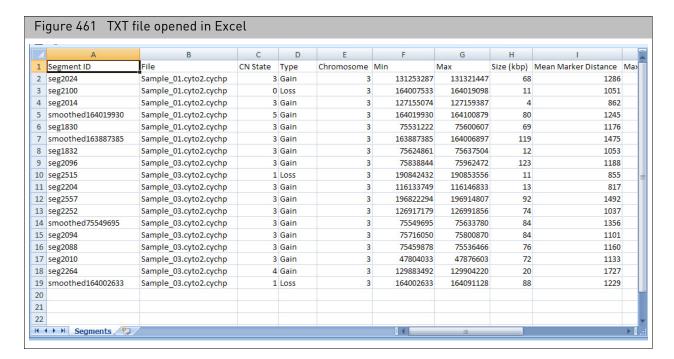
2. In the Table tool bar, click the Export **TXT** button. The Save as TXT window opens. (Figure 460)



- 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 (Figure 461).

It can be opened using a text editing or spreadsheet program, or in other software designed to use Tab Separated Value TXT format.



Exporting a segments table with modified segments to a TXT file

In the Segments Table:

- Materially Modified segments are shown in italic text when Edit Mode is ON, but not italicized when Edit Mode is OFF.
- Calls and Interpretations don't cause a segment's row text to be italicized in Edit Mode.
- Deleted segments are shown in strike-through text when Edit Mode is ON, and are not present in the table when Edit Mode is OFF.
- In the TXT Exports of the Segments Table table, please note that Deleted segments will be part of the export when Edit Mode is ON, and will not be part of the export if Edit Mode is OFF

Figure 462 is an example of a Segment Table with Edit Mode **ON**. Note its italicized text representing materially modified segments and the strike-through text showing deletions.

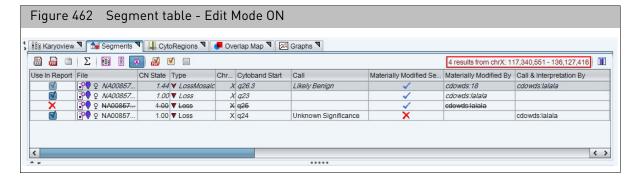


Figure 463 is an example of a Segment table that has been exported with the Edit Mode

**ON**. Note the 4th row of the Use in Report column. In the case of this segment, it reads FALSE, because this segment was deleted.

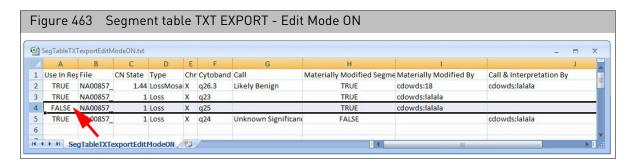


Figure 464 is an example of a Segment Table with Edit Mode **OFF**. The italicized text representing materially modified segments is no longer present. The deleted segment and strike-through text showing a deletion (shown in Figure 462 and Figure 463) also do not appear.

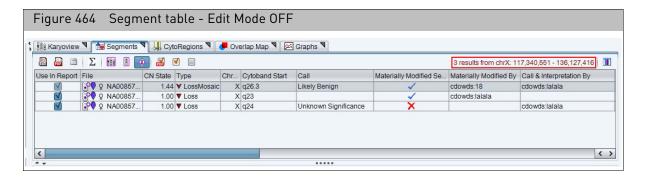
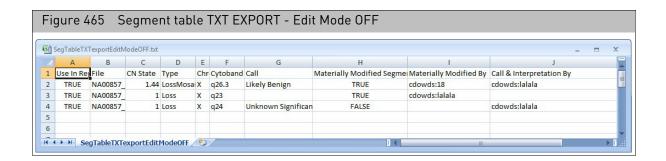


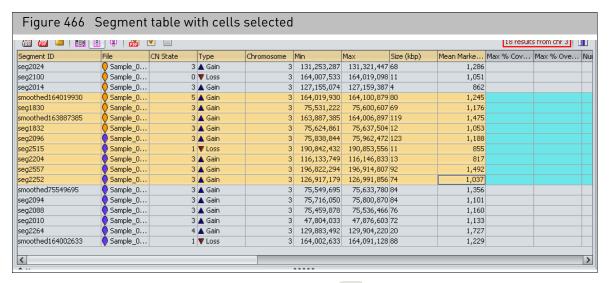
Figure 465 is an example of how an exported TXT table appears with Edit Mode **OFF**. Note the deleted segment shown in Figure 462 and Figure 463 is not present.



## Transfer to clipboard

You can copy data from selected cells to the clipboard for pasting into a text or spreadsheet file.

1. Select the cells you want to copy in the table. (Figure 466)



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 (Figure 467).

Figure 467 Data pasted into text file									
smoothed164019930   Sample_01.cyto2.cychp   seg1830   Sample_01.cyto2.cychp   3.0   Gain   smoothed163887385   Sample_01.cyto2.cychp   3.0   Gain   seg2396   Sample_03.cyto2.cychp   3.0   Gain   seg2515   Sample_03.cyto2.cychp   3.0   Gain   seg2515   Sample_03.cyto2.cychp   3.0   Gain   seg2557   Sample_03.cyto2.cychp   3.0   Gain   seg2557   Sample_03.cyto2.cychp   3.0   Gain   seg2252   Sample_03.cyto2.cychp   3.0   Gain   seg2252   Sample_03.cyto2.cychp   3.0   Gain   seg2252   Sample_03.cyto2.cychp   3.0   Gain   seg2252   Sample_03.cyto2.cychp   3.0   Gain   Sample_03.cyto2.cychp   3.0	5.0 33.0 33.3 33.3	Gain 3 75531222 Gain 3 75624861 75838844 190842432 116133749 196822294 126917179	164019930 75600607 163887385 75637504 75962472 190853556 116146833 196914807 126991856	164100879 69 1176 164006897 12 1053 123 1188 11 855 13 817 92 1492 74 1037	80 119	1245 1475	<		

## Combining PDFs into a single PDF

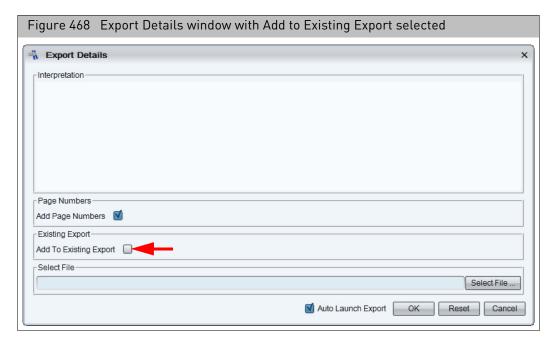
You can do this by:

- Adding new data to an existing PDF file.
- Merging two or more existing PDF files.

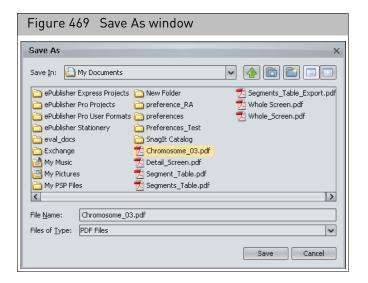
**IMPORTANT!** The results from ChAS are for Research Use Only. Not for use in diagnostic procedures.

# Adding a new PDF export to an existing PDF file

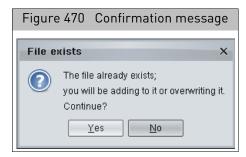
1. When exporting a table or graph as a PDF, click the **Add to Existing Export** check box in the **Export Details** window. (Figure 468)



2. Click **Select File** and select an existing PDF file for the data to be added to. The **Save As** window opens. (Figure 469)



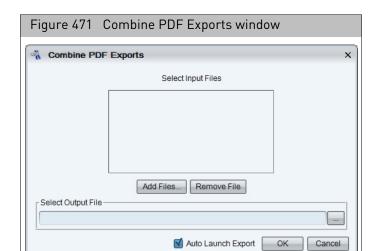
Select a PDF file, then click Save in the Save As window.
 A confirmation message appears, asking if you want to overwrite or add to the data (Figure 470).



- 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 **Export Details** window. The new report data is combined with the existing report.

## Combining two existing PDF files

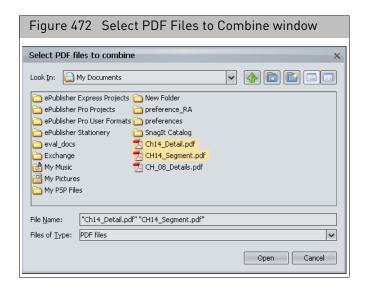
1. Click Exports  $\rightarrow$  Combine PDFs.



The Combine PDF Exports window opens. (Figure 471)

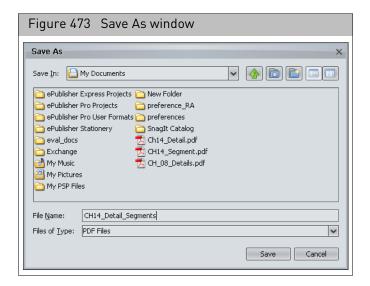
2. Click Add Files....

The Select PDF Files to combine window opens. (Figure 472)



- 3. Select the PDF files to combine, then click **Open** in the Select PDF Files window. 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 window opens. (Figure 473)



5. Enter a file name for the combined PDF file, then click **Save** in the **Save As** window.

You are returned to the **Combine PDF Exports** window.

6. Click OK.

Your selected PDFs are combined.

## Exporting with ClinVar

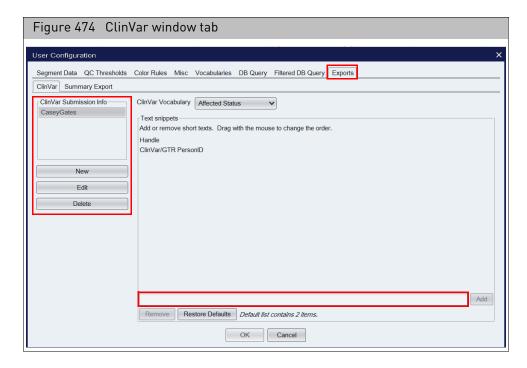
The ClinVar export enables exporting of copy number data using the ClinVar Full Template for easy submission directly to ClinVar. A ClinVar Submission Profile must be created to use this Export. For details on the template, go to: www.ncbi.nlm.nih.gov/clinvar/docs/submit/

**IMPORTANT!** ChAS is a research use only application and any submission to ClinVar is the responsibility of the submitter.

All fields exported into the ClinVar submission template are selected and defined by the user. Standard ClinVar nomenclature is provided for required submission fields and can be customized as shown in "Adding vocabulary content" on page 433.

## Creating a ClinVar submission profile

- Click Preferences → Edit User Configuration.
   The User Configuration window appears.
- 2. Select the **Exports** tab, then select the **ClinVar** tab. (Figure 474)



- 3. In ClinVar Submission Info pane, click the **New** button to create a new submission profile.
  - An Edit Profile window appears.
- 4. Complete all the appropriate fields. Fields with an \* are required by ClinVar.
- 5. Click **OK** to save the profile.
- 6. Optional: To create additional ClinVar profiles, repeat steps 3-5

# Editing an existing ClinVar profile

- 1. Highlight the profile name in the ClinVar Submission Info list box that is to be modified.
- 2. In ClinVar Submission Info pane, click the Edit button.
- 3. Edit the appropriate fields.
- 4. Click **OK** to save your changes.

#### Deleting a profile

- 1. Highlight the profile name in the ClinVar Submission Info list box you want to delete.
- 2. Click the **Delete** button.
- 3. Click **OK** to confirm the profile deletion, or click **Cancel** to keep the profile.

# Adding vocabulary content

By default, recommended ClinVar vocabularies are stored for certain required fields, but additional terms can be added to any field.

- 1. Click Preferences  $\rightarrow$  Edit User Configuration.
  - The User Configuration window appears.
- 2. Click the Exports tab, then click the ClinVar tab.
- 3. Select a category you want to add a term(s) to, then use the text field (Figure 474) to enter the additional term(s).
- 4. From the ClinVar vocabularies drop-down, select the category that you want to add a term(s) to.
- 5. Use the text field (at the bottom) to enter the additional term(s).
- 6. Click the **Add** button to add the term to the category's list.

# Removing vocabulary content

- 1. From the ClinVar vocabularies drop-down list, select a category that contains the term you want to remove.
- 2. Highlight the term, then click **Remove**.

### Exporting in ClinVar format

There are certain fields that are required before uploading to ClinVar. It is recommended you use the **ClinVar Table State** in the Segments Table to expose all required fields. To use the ClinVar Table State, refer to "Saved table states" on page 332.

- 1. In the Segments Table, apply the ClinVar Table State.
- 2. Select the segments to be exported using the **Use in Export** check box.
- 3. Fill in all columns for the selected segments, as all columns in the ClinVar Table State are required before you can upload to Clinvar.
- 4. In the Exports Menu, select ClinVar Export.
  - A browse window appears.
- 5. Select a location to save the export, then use the File Name text box to name the export, then click **OK**.
  - A Submission Info/Segment Data window appears.
- 6. Select the ClinVar Profile you want to use for this export.
- 7. Optional: If you want to add any additional comments to the export, click on the Segment Data tab to enter them within this tab.
- 8. Click **OK** to export.

**Note:** You can open the ClinVar export in Excel to add information to the optional fields. Opening the ClinVar export from ChAS, auto-populates all currently required ClinVar fields.

Table 19 Variant tab: Auto-populated columns into the ClinVar submission template (all other optional columns are blank upon export).

Auto-populated column	Description	
Chromosome	Populated from Segment Table (Chromosome)	
Variant type	Populated from Segment Table (Gain/Loss)	
Variant length	Populated by ClinVar upon submission	
Copy Number	Populated from Segment Table (Gain/Loss)	
Variation identifiers	Populated using OMIM track	
Condition category	Populated from ClinVar Vocabulary	
Clinical significance	Populated from Segment Table (Call)	
Date last evaluated	Uses date of export unless otherwise specified on the Submission Info window before exporting.	
Comment on clinical significance	Populated from Segment Table (Segment Interpretation)	
Collection method	Populated from ClinVar Vocabulary	
Allele origin	Populated from ClinVar Vocabulary	
Affected status	Populated from ClinVar Vocabulary	
Structural variant method/ analysis type	SNP Array	
Platform type	Microarray	
Platform name	Populated based on the Microarray used	
Software name and version	Chromosome Analysis Suite 4.3	

Table 20 Case Data tab: Auto-populated columns into the ClinVar submission template (all other optional columns are blank upon export).

Auto-populated column	Description
##Linking ID	Populated from Segment Table (Full Location)
Collection method	Populated from ClinVar Vocabulary
Allele origin	Populated from ClinVar Vocabulary
Affected status	Populated from ClinVar Vocabulary
Structural variant method/ analysis type	SNP Array

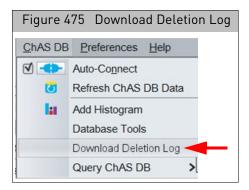
Table 20 Case Data tab: Auto-populated columns into the ClinVar submission template (all other optional columns are blank upon export).

Auto-populated column	Description
Clinical Features	Populated from Sample Properties (Phenotype)
Tissue	Populated from Sample Properties (Sample Type)
Sex	Populated from Gender determination
Platform type	Microarray
Platform name	Populated based on the Microarray used
Software name and version	Chromosome Analysis Suite 4.3

### Exporting a deletion log

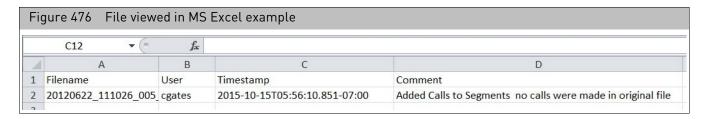
The user, time, filename and reason why a file is deleted or republished to a ChAS database can be exported.

1. Click ChAS DB → Download Deletion Log (Figure 475)



Your default Internet browser opens.

- 2. Enter your ChAS Database credentials, as you normally would.
- 3. Depending on your Internet browser, you may be prompted to either download the file or save it. To save it, click **File**  $\rightarrow$  **Save As.**
- 4. To view your exported log file, open it in Excel. (Figure 476)



Deletion Logs can also be exported from the ChAS DB Tools Maintenance Page. See "Downloading deletion logs" on page 452.



# User profiles and named settings

ChAS provides many options for customizing the display of data and annotations.

The User Profiles and Named Settings functions enable you to save your analysis and display settings and are described in the following sections:

- "Types of settings"
- "Creating and using user profiles" on page 438
- "Creating and using named settings" on page 439
- "Exporting and importing preferences" on page 442

### Types of settings

ChAS provides two ways to store setup information. Each way works differently and performs different functions.

- "User profiles"
- "Named settings" on page 437

#### User profiles

A ChAS Browser 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 as described in the table below. These named settings cannot be deleted.

Named Setting	Genome Segment Filters	CytoRegion Segment Filters	Data Types
Standard	Gain: Marker Count = 50 Size (kbp) = 400; Loss: Marker Count = 50 Size (kbp) = 400	Gain: Marker Count = 50 Size (kbp) = 400 Loss: Marker Count = 50 Size (kbp) = 400	Gain Loss GainMosaic LossMosaic Copy Number State Weighted Log2 Ratio Allele Peaks Allele Difference
High Resolution	Gain: Marker Count = 50 Size (kbp) = 100; Loss: Marker Count = 50 Size (kbp) = 100	Gain: Marker Count = 25 Size (kbp) = 50 Loss: Marker Count = 25 Size (kbp) = 50	Gain Loss GainMosaic LossMosaic Copy Number State Weighted Log2 Ratio
LOH only (3Mb and 50 SNPs)	LOH: Marker Count = 50 Size (kbp) = 3000	LOH: Marker Count = 50 Size (kbp) = 3000	LOH Genotype Calls Allele Peaks Allele Difference
Differential Gains and Losses	Gain: Marker Count = 50 Size (kbp) = 400; Loss: Marker Count = 50 Size (kbp) = 100	Gain: Marker Count = 25 Size (kbp) = 50 Loss: Marker Count = 25 Size (kbp) = 50	Gain Loss GainMosaic LossMosaic Copy Number State Weighted Log2 Ratio

Named Setting	Genome Segment Filters	CytoRegion Segment Filters	Data Types
OncoScan Defaults	0	0	0
XON-Level 1	XON Level: Level 1 = On Level 2-4 = Off XON Gain/Loss Marker: Count = 0 Size (kbp) = 0	0	XON Region Gain, XON Region Loss, LOH Segments, Log2 Ratio, Smooth Signal, Allelic Difference

### Creating and using user profiles

You can only select or create user profiles upon starting ChAS.

Double-click on the ChAS icon on the desktop; or
 From the Windows Start Menu, select Programs → Thermo Fisher Scientific → Chromosome Analysis Suite → Chromosome Analysis Suite.

The ChAS Splash Screen and the **Select User** window open. (Figure 477)

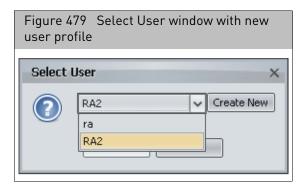


Click Create New in the Select User window.The Create New User window opens. (Figure 478)



- 3. Enter a name for the new profile in the User ID field.
- 4. Click **OK** in the Create New User window.

The new user appears in the drop-down User list in the **Select User** window. (Figure 479)



5. Select the new user, then click **OK**.

Any changes you make to the setup of the software that is listed in "User profiles" on page 436 will be saved when you shut down the software and used the next time the software is opened with this user profile.

# Deleting a user profile

1. Go to the Windows folder where the user profiles are stored and delete the folder with the profile name you want to delete.

You can see the location of the folder in the About window, as described in "Analysis file locations in Windows 10" on page 26.

#### 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:

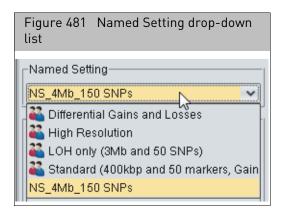
- Which graphs and segment types are turned on or off
- Segment Filter Settings
- Restricted mode on/off
- 2. From the Preferences menu, select **Save Named Setting**.

The Setting Name window opens. (Figure 480)



3. Enter a name for the setting you want to create, then click **OK**.

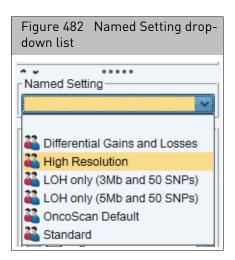
The setting is saved and appears in the Named Setting drop-down list. (Figure 481)



**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.

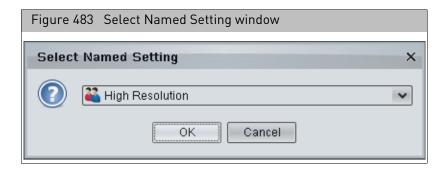
# Selecting a named setting

1. From the Named Setting drop-down list, select the setting. (Figure 482)



Alternatively, from the Preference menu, select Apply Named Setting...

#### The Select Named Setting window opens. (Figure 483)



2. Select the Named Setting from the drop-down list, then 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:

From the Preferences menu, select **Delete Named Setting**.
 The **Delete Setting** window opens. (Figure 484) **Note:** Shared Named Settings (the icon in the Named Setting list) do not appear in the Delete Setting dropdown list. Users cannot delete or modify a shared Named Setting.



2. Select the setting you want to delete from the drop-down list, then click **OK**. The setting is deleted.

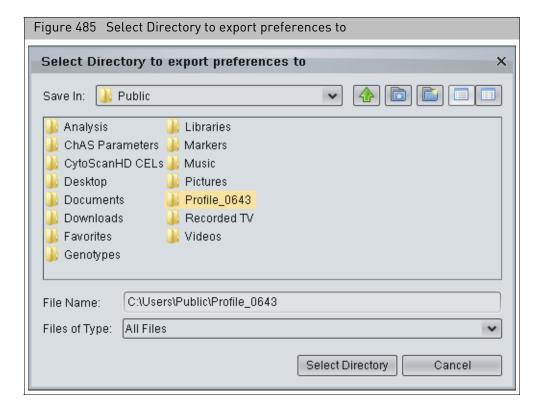
#### Exporting and importing preferences

Preferences functions enable 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.

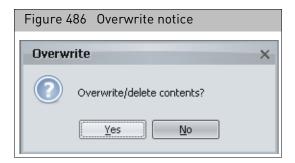
# Exporting preferences

From the Preferences menu, select Export Preferences...
 The Select Directory to export preferences to window opens. (Figure 485)



- 2. Use the navigation features of the window to select or create a directory for the preferences. **Note:** The software creates a folder named "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
- 3. Click **Select Directory**.

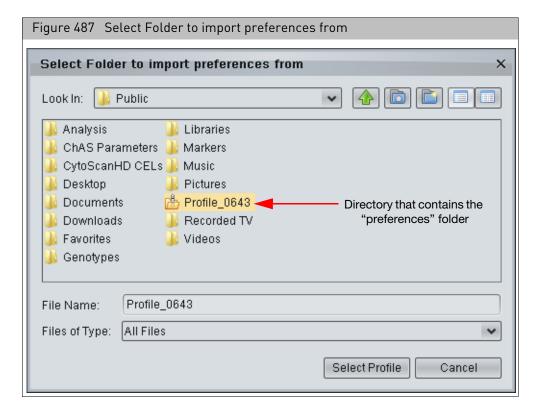
If the directory already contains a "preferences" folder, the Overwrite notice appears. (Figure 486)



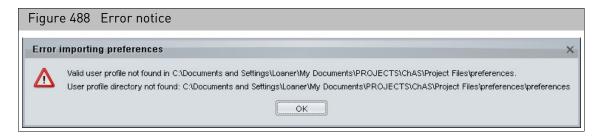
4. Click **Yes** to export the preference files to the directory that you selected. You can then transfer the preferences to another user profile or system.

# Importing preferences

From the Preferences menu, select Import Preferences...
 The Select Folder to import preferences from window opens. (Figure 487)



- 2. Use the navigation features of the window 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 488)



Click **OK** and repeat steps 1 through 3, selecting the correct folder.

The Restart notice appears. (Figure 489)



**Note:** The imported preferences will not be applied until you restart ChAS.

# Importing hyperlinks

External websites may update their links from time to time. To remedy this, a feature as been added to update an outdated link(s) within ChAS.

1. Click Preferences  $\rightarrow$  Import Hyperlinks.

The Load Hyperlinks Configuration window opens.

- 2. Navigate to your updated hyperlinks (.chaslink) file, click to highlight it, then click Select File.
- 3. Restart the ChAS Browser to apply the link update(s).

**Note:** For .chaslink file help, contact Technical Support.

# Configuring the HTTP service

The HTTP service may enable external applications to interact with the ChAS Browser. By default, this service is off. Please contact Technical Support before activating this feature.

1. Click Preferences → Configure HTTP Service.

The Configure HTTP Service window opens.

- 2. Click the check box to enable the service, then enter the designated Port number.
- 3. Close the window.

The service and Port are now activated.



### Database tools

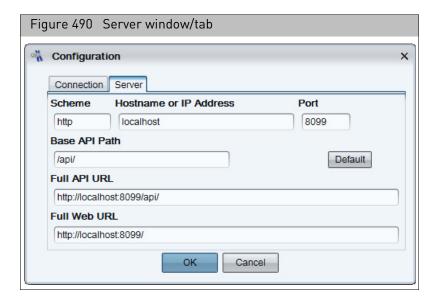
### Connecting to a remote ChAS DB server

The ChAS v4.3 browser cannot point to a ChAS v3.0/3.1/3.2/3.3/4.0/4.1/4.2/4.2.1 database.

You must upgrade the ChAS v3.0/3.1/3.2/3.3/4.0/4.1/4.2/4.2.1 database to ChAS v4.3 through the Backup and Restore process. See "Backing up a database" on page 449 and "Restoring a database" on page 450.

- From the Preferences drop-down menu, click Edit Application Configuration...
   The Configuration window appears.
- Click the Server tab.The Server window/tab appears. (Figure 490)
- 3. Type in the name or IP address of the computer/server name that you would like to connect to in the Hostname or IP Address text field or contact your IT Department for help completing this form.

**Note:** Up to three Hostnames/IP Addresses can be saved/stored.



4. Click OK.

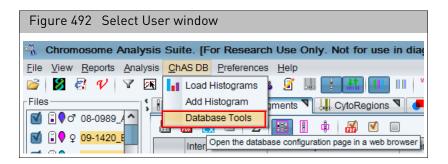
If a connection to the ChAS DB cannot be established or the server/computer containing ChAS DB is not turned on, the following message appears (Figure 491). Please check the name/IP address and make sure the server/computer is turned on.



#### Accessing the ChAS DB server tools

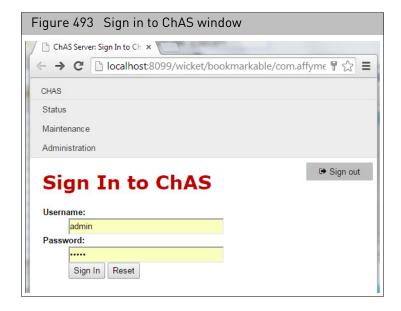
**Note:** The screen captures used in this chapter may vary depending on which Browser you are using.

1. From the Chas DB drop-down menu, click to select **Database Tools**. (Figure 493)



The following web page appears. (Figure 493)

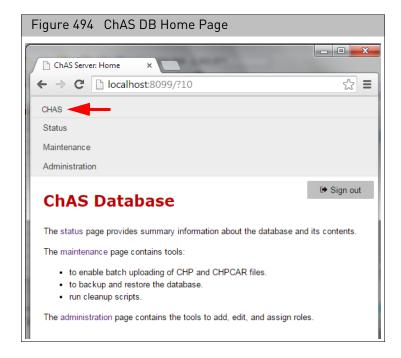
**IMPORTANT!** The ChAS Server Home Page requires an active Internet connection, requires a browser (Chrome and Internet Explorer v11 are recommended). Also, if you are using the local ChAS DB, an active Internet connection is not required.



2. Log in using the installer's factory default Username: **admin** and Password: **admin**. After logging in, it is recommended new users go to "Administration" on page 454 to create a New User(s) and/or edit User(s) roles.

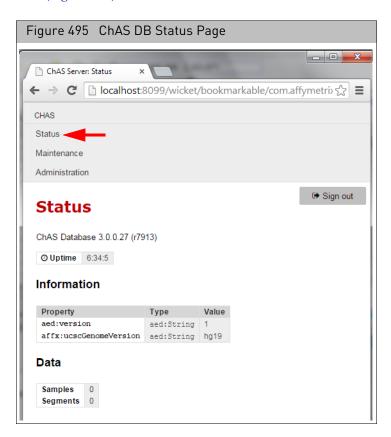
**Note:** Make sure you look in the URL field to identify which ChAS database the ChAS Database Tools is accessing.

The ChAS DB Home Page appears. (Figure 494)



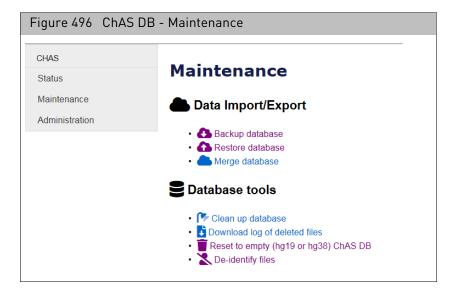
### Status page

Use this page to view how many samples and segments are in the Database. (Figure 495)



#### Maintenance

Use this page to perform a backup, restore, and database clean up. (Figure 496)



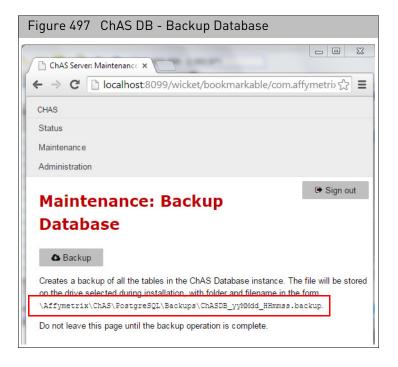
### Backing up a database

1. Click **Backup Database**, then click the **Backup** button.

A backup is automatically generated and is stored in the Affymetrix directory shown in Figure 497.

(\Affymetrix\ChAS\PostgreSQL\Backups\ChASDB\_yyMMdd\_HHmmss.backup) **Note:** A backup is automatically done whenever a Restore operation is performed.

**IMPORTANT!** It is strongly recommended that you perform scheduled routine backups of the database.



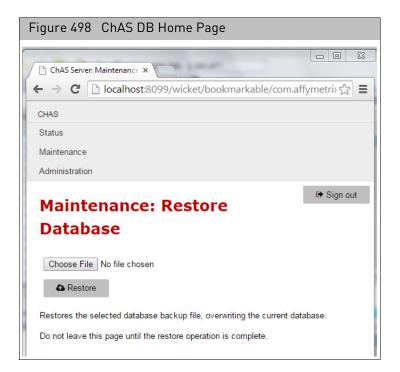
**Note:** The ChAS installer creates a disabled Windows Task that automatically backs up the ChASDB database on a weekly basis once it is enabled.

### Restoring a database

**Note:** Restoring a backup file created in ChAS 3.0/3.1/3.2/3.3/4.0 automatically upgrades it to ChAS 4.1.

- 1. Click **Restore Database**, then click the **Choose File** button. (Figure 498) An Explorer window opens.
- Navigate to the location where your ChAS DB was last backed up, then click Open.
   By default, a backup of your current database is stored/resides here: \\Affymetrix\ChAS\\PostgreSQL\\Backups
- 3. Click **Restore** to start the restore process.

**IMPORTANT!** Do not leave this page once choosing the Restore button until you see the message that the database has been successfully restored. Also, once the restore process has successfully completed, you must click ChAS DB  $\rightarrow$  Refresh ChAS DB Data to view the data in the database using the ChAS Browser.



**IMPORTANT!** After restoring the database, you must click ChAS DB  $\rightarrow$  Refresh ChAS DB data to view the newly restored data from the database.

#### Merging ChAS databases

When merging the contents of two ChAS databases into a single database, one database should be restored into ChAS and the other database must exist as a **backup.db** file.

When merging the segments from two databases, if a duplicate entry is found then the merge keeps the entry for the database currently active in ChAS. The duplicate from the **backup.db** is skipped.

### Merging two ChAS databases

**IMPORTANT!** The two ChAS databases to be merged, must be from the same version of ChAS. Also, the library files for CytoScan HD and OncoScan CNV Plus must be present in your Library folder before merging the two databases.

Make sure one of the ChAS databases is restored into ChAS (for details on how to Restore a ChAS DB, see page "Restoring a database" on page 450). Since any duplicate segments between the databases will keep the copy from the actively restored database, make sure the database with the more complete content is the one that is actively restored in ChAS.

- From the ChAS browser, go to ChAS DB → Database Tools.
   If prompted, log into the ChAS database as you normally would.
- 2. Click on the **Maintenance** link.
- 3. Click on the **Merge database** link.
- 4. Click the **Browse** button to navigate to the **backup.db** file you want to merge with the current database.
- 5. Click Merge.

# Merging an older ChAS database

If you want to merge a database (from an older version of ChAS) with a current ChAS 4.0 database, perform these steps:

- 1. Backup your current ChAS 4.0 database.
- 2. Restore the database from (e.g.) ChAS 3.1. See "Restoring a database" on page 450. During the restore process, the older ChAS database is automatically upgraded and is now compatible with ChAS 4.0.
- 3. Backup the 3.1 database you just restored.
- 4. Restore the ChAS 4.0 database you backed up in Step 1.

**Note:** If the databases to be merged contain duplicate entries, the copy that is in the currently restored database will be kept. The entry from the backup.db that is being merged will be skipped.

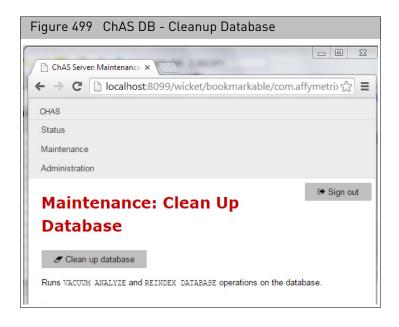
- From the ChAS browser, go to ChAS DB → Database Tools.
   If prompted, log into the ChAS database as you normally would.
- 6. Click on the Maintenance link.
- 7. Click on the **Merge database** link.
- 8. Click the **Browse** button to navigate to the **backup.db** file you want to merge with the current database.
- 9. Click Merge.

### Cleaning up a database

ChAS DB will automatically run re-indexing scripts to maintain optimal performance. You can also run these scripts manually if desired.

**Note:** You must have a **Manager** or an **Admin** role to clean up a database.

 Click the Clean up database button (Figure 499) to run the Vacuum Analyze and Reindex Database optimization process.



### Downloading deletion logs

Use this feature to export a list of filenames that were deleted from ChAS DB.

- 1. Click the Download Log of Deleted Files
- 2. If asked, provide your ChAS DB Username and Password.
- 3. Save the genome-model log file.
- 4. Open with Microsoft Excel for easier viewing.

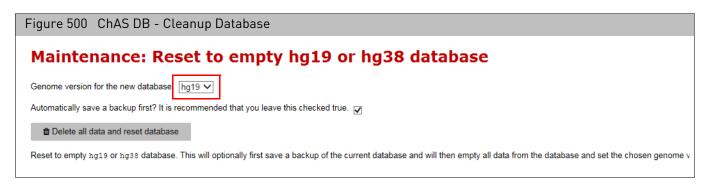
### Creating a blank ChAS DB

Use this feature to create a blank hg19 or hg38 ChAS DB. Make sure you backup your previous database prior to creating an empty ChAS DB, as the database will be erased and recreated.

**Note:** xxCHP files are only compatible with a ChAS DB of the same genome version.

- 1. Back up the current ChAS DB.
- 2. Click on the Reset to Empty (hg19 or hg38) ChAS DB

3. Use the drop-down to select the genome version for your new database. (Figure 500)



- 4. Check the box to have a backup of the current ChAS DB before the database is deleted and an empty DB is created.
- 5. Click Delete all data and reset database.

### **Deidentifying Files**

Deidentifying files will remove potentially sensitive information from the ChAS DB. By running Deidentification, the file names stored in the ChAS DB will be replaced by an alpha numeric ID.

**Note:** If you have included sensitive information in custom database fields, this process will not remove that information.

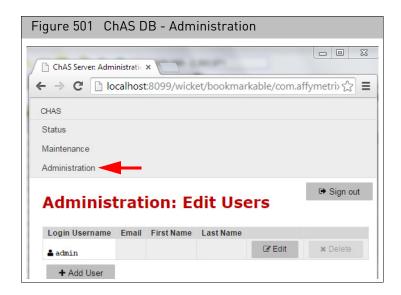
- 1. Click on De-identify Files.
- 2. (Optional) Create a backup of the database to preserve the original content, then click the **Start De-identification** button to replace the file names in your ChAS Database.

#### **Administration**

**Note:** You must have an Admin role to perform Administration functions. Log files for the ChAS database can be found in: \ProgramData\Affymetrix\ChAS\Log

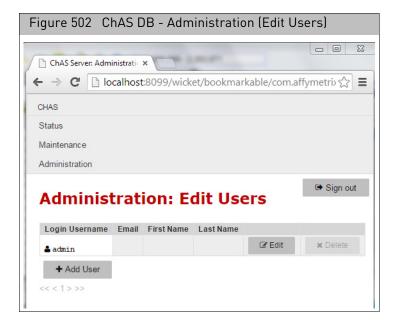
#### 1. Click Administration.

The following window appears: (Figure 501)



#### 2. Click Edit Users.

The following window appears: (Figure 502)



Do one of the following:

- Click the **Edit** button to edit a current Username.
- Click **Add User** to add (grant privileges to) a new ChAS DB user.

#### **Permission Guidelines**

- **User Role** has permission to query the database, edit segments, add segment and sample annotations, but does not have permission to upload data to the database, run backup/restore, maintenance or edit users.
- Manager Role has permission to query the database, edit segments, add segment
  and sample annotations, upload data to the database, run backup/restore,
  maintenance, but does not have permission to edit users.
- Admin Role has permissions to run all functions in ChAS Browser and ChAS DB.

Using a shared ChAS database while off-line

**IMPORTANT!** If your Windows Firewall is enabled during the installation of ChAS and you want to Backup the ChAS Database and Restore it to your local ChAS DB, a message may appear indicating that you cannot connect to the shared folder. If this message appears, contact your IT department for help in allowing file sharing through the Windows Firewall.

#### Working off-line using a ChAS database that resides on a shared server

- Connect to the shared ChAS DB server click Preferences → Edit Application
   Configuration, ChAS DB server tab. For details on connecting to another ChAS
   DB, see "Connecting to a remote ChAS DB server" on page 445.
- 2. Follow the procedure to perform a back up, locate the back up, then copy it to your local computer.
- 3. Return to **Preferences** → **Edit Application Configuration**, **ChAS DB server** tab, then click on the default to redirect to your local ChAS DB server.

See "Restoring a database" on page 450 for instructions on how to restore the back up database - you just copied from the server.

**IMPORTANT!** You must have a Manager or Admin role and make sure you log back into the local host before restoring your computer. While performing a restore from a backup of a shared server, the roles associated with the shared server are displayed, therefore any roles that were created on the local server are replaced by those used on the shared server (until local host ChAS DB is restored again).

### Publishing data you have analyzed in off-line mode to the shared ChAS DB server

- Log back into the shared ChAS DB server. To do this. click Preferences → Edit
   Application Configuration, ChAS DB server tab, then enter the server name/IP
   address.
- 2. Click OK.

**Note:** After you are logged into the shared ChAS DB server, publish the samples to the database as you normally would. See "Publishing data to the database" on page 404. If a xxCHP file has been previously Published to the database, you will receive a warning indicating this sample already exists in the database. You can choose to overwrite the existing information or cancel to keep the existing information.

### Remapping a hg19 ChAS DB to hg38

A ChAS DB populated with data from hg19 analyses can be remapped to hg38 coordinates using the ChAS DB Remapper feature.

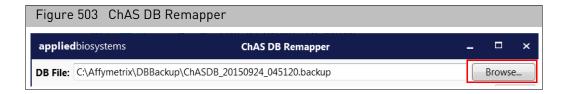
The ChAS DB Remapper maps the segments from your hg19 ChAS DB to the hg38 genome. All segments in ChAS DB start and end with a marker currently mapped in hg19. The Remapper takes the probe locations for all markers in the segment from hg19 and remaps them to hg38. It locates the best representation of the remapped segment in hg38 by taking into account the number of additional markers in the remapped segment, as well as the number of markers removed from the original segment.

Remap confidence = (originalMarkerCount - addedCount - 2\*removedCount)/ originalMarkerCount

For remapped segment definitions in a ChAS DB, see the table on page 393.

Segments with a Remap confidence >0.75 are remapped to the hg38 ChAS DB. The Remapper makes a copy of the database prior to remapping. However, always make a backup or copy of the ChAS DB that you would like to remap for safe keeping.

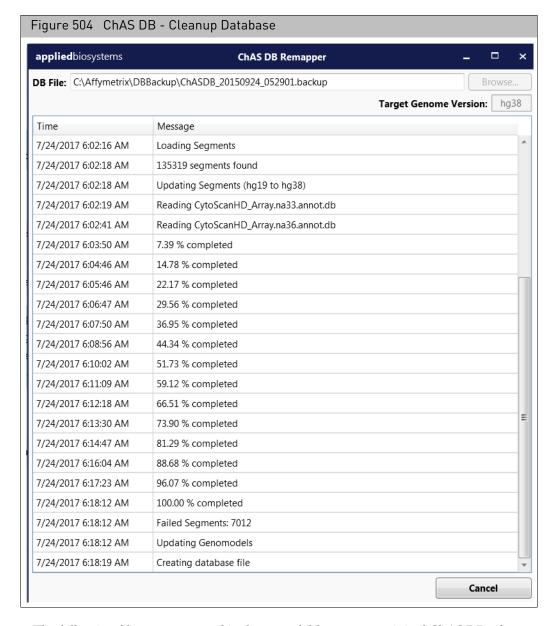
- 1. Click on ChAS DB  $\rightarrow$  ChAS DB Remapper.
- 2. Click the **Browse** button (Figure 503) to select a ChAS db.backup to be remapped to hg38.



#### Click Start.

**Note:** Depending on the size of the ChAS DB.backup to be remapped, this process may take several minutes.

The ChAS Remapper window appears. (Figure 504)



The following files are generated in the same folder as your original ChAS DB select to remap:

- ChAS DB.hg38.backup this backup can be restored as the active ChAS DB for querying within the Browser.
- A TXT file listing all original segments provides a text file of the original segment information and the remapped segment information along with success or fail criteria.
- A TXT file listing the segments that failed to remap to hg38 provides s text file of the original segments that did not remap to the new genome build.

For more information on viewing a remapped database, see "Additional segment intersection information" on page 395.



### ChAS Database Loader (CDL)

CDL is now part of the ChAS Browser. CDL enables uploading of xxCHP files from any previous (including the current) version of ChAS. You may upload up to 500 xxCHP files at one time (as long as they are all analyzed from the same genome build).

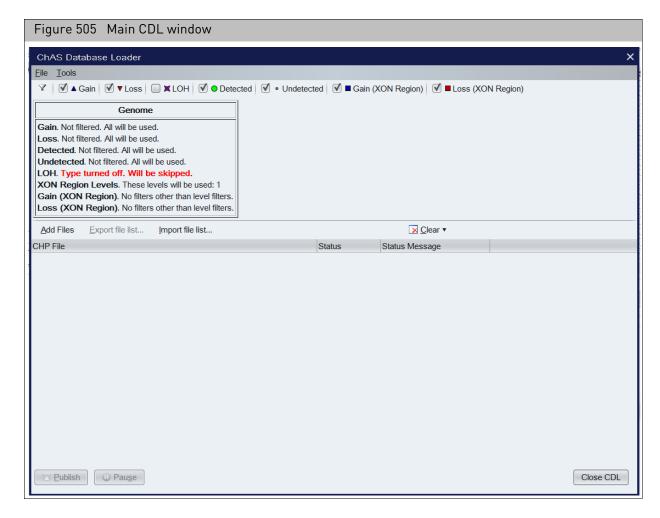
CDL supports the following array types:

- Genomewide SNP 6
- CytoScan 750K
- CytoScan HD
- CytoScan Optima
- CytoScan XON
- CytoScan HTCMA
- OncoScan CNV Plus
- OncoScan CNV
- Cytogenetics 2.7M

### **Starting CDL**

1. Click ChASDB → ChAS Database Loader

The ChAS Database Loader window appears. (Figure 505)

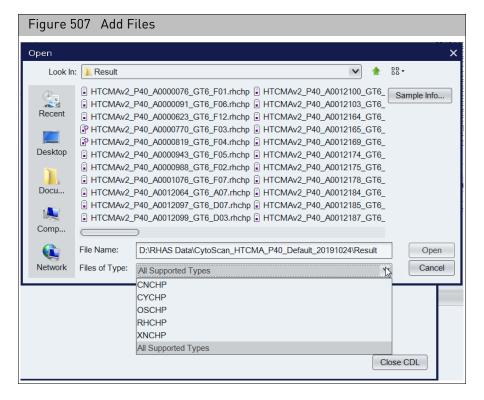


### Adding files to CDL

1. From the main CDL window, click Files  $\rightarrow$  Add Files (Figure 506)



An Explorer window appears. (Figure 507)



By default, the **Files of Type** is set to All Supported Types. (Figure 506) If you want to view a specific supported file type, click the drop-down, then select the file extension you want to display.

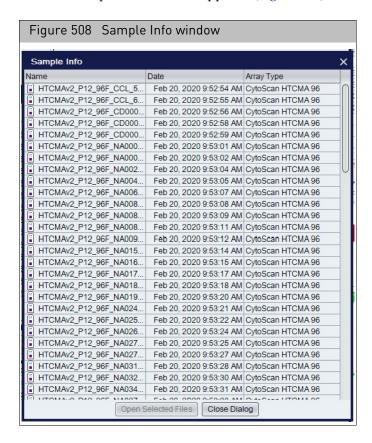
- 2. Single click, Ctrl click, Shift click or Ctrl a (to select multiple files).
- 3. Click Open.

Your selected files now appear in CDL's main window. (Figure 509)
Repeat steps **1-3** if you want to add files (up to 500) from different saved locations.

#### Sample info

1. If you want to view the properties of the files displayed in the Explorer window, click Edit Configuration.

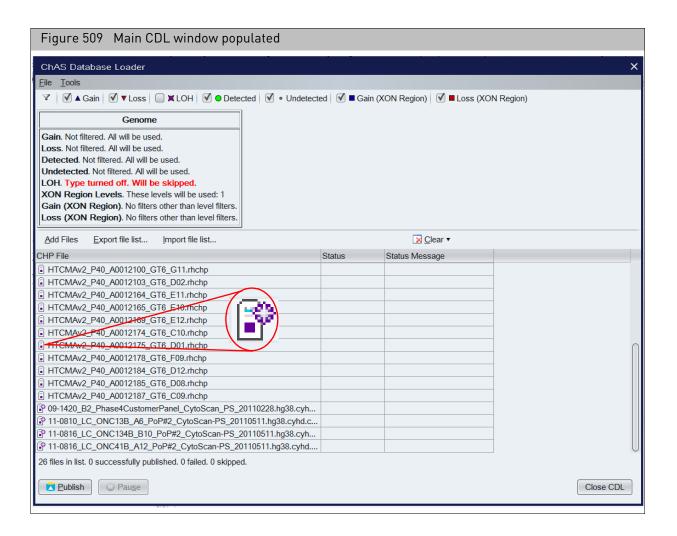
The Sample Info window appears. (Figure 508)



- 2. Single click, Ctrl click, Shift click or Ctrl a (to select multiple files).
- 3. Click Open Selected Files.

Your selected files now appear in CDL's main window.

**Note:** A special icon is used to indicate a CHPCAR or "sidecar" file, as shown in Figure 509. For more information on sidecar files, "Editing segment data overview" on page 225.



IMPORTANT! File level properties are optional fields that are available to CHP files analyzed in ChAS 3.0 or higher. Any file level properties entered are stored in the CHPCAR file, these properties will not populate in the CDL window. However, if those properties were entered for a CHP file and are contained in the CHPCAR file, they will be published to the database. Entries in a CHPCAR file supersede entries in the text file. File level properties for your xxCHP files can be added directly to the database after publishing has completed. For more details, see "Interacting with the ChAS database" on page 387.

# Adding files to be published using a text file

Files can be loaded into CDL using a tab-delimited text file. Place the file names, including their paths in the first column, then label the first column header **CHP File** as shown in Figure 510.

F	igure 510 Tab-delimited text file Header 1 example
	A
1	CHP File
2	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0000076_GT6_F01.rhchp
3	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0000091_GT6_F06.rhchp
4	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0000623_GT6_F12.rhchp
5	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0000770_GT6_F03.rhchp
6	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0000819_GT6_F04.rhchp
7	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0000943_GT6_F05.rhchp
8	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0000988_GT6_F02.rhchp
9	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0001076_GT6_F07.rhchp
10	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0012064_GT6_A07.rhchp
11	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0012097_GT6_D07.rhchp
12	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0012099_GT6_D03.rhchp
13	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0012100_GT6_G11.rhchp
14	D:\RHAS Data\CytoScan_HTCMA_P40_Default_20191024\Result\HTCMAv2_P40_A0012103_GT6_D02.rhchp
15	D:\RHAS Data\CvtoScan HTCMA P40 Default 20191024\Result\HTCMAv2 P40 A0012164 GT6 F11.rhchp

- Click on Import File List or click Files → Import file list.
   An Explorer window appears.
- 2. Navigate to, then select the tab-delimited text file containing path to the xxCHP files to be loaded into CDL.

#### Publishing to the ChAS database

**IMPORTANT!** Before you use CDL to publish your files, it is highly recommended you backup your ChAS database first. For instructions on how to access and backup your database, refer to Chapter 21, "Database tools" on page 445. Also, xxCHP files can ONLY be published to a ChAS DB of the same genome version assignment.

# Testing your connection (optional)

Before publishing, you may want to test your ChAS database connection. To do this:

- 1. Click ChAS DB → ChAS Database Loader
- 2. Click **Tools**  $\rightarrow$  **Test Connection**

A message appears if there is a successful connection to the ChAS database.

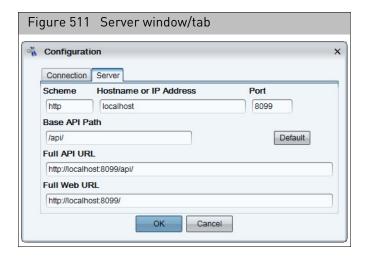
# Verifying the ChAS database

1. From the ChAS Browser, click the **Preferences** drop-down menu, then select **Edit Application Configuration...** 

The **Configuration** window appears.

2. Click the **Server** tab.

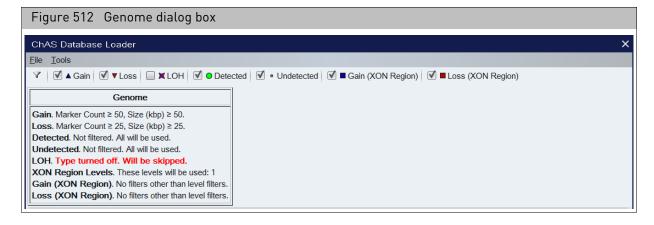
The Server window/tab appears. (Figure 511)



3. Verify the ChAS DB you are publishing to is correct, then click **OK**.

## Before publishing files

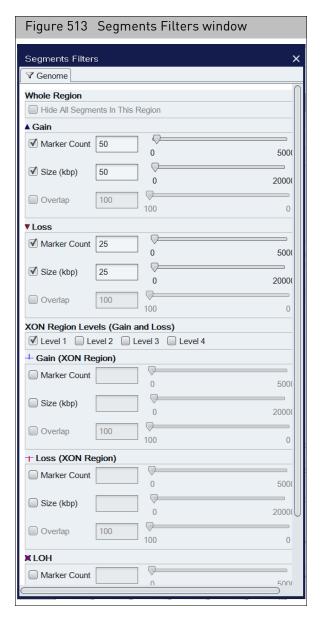
Before publishing files, you must check the Genome dialog box (Figure 512) to make sure your desired filter settings and data types are enabled.



**Note:** QC thresholds and Smooth/Joining settings in the ChAS Browser will be used when publishing xxCHP files using CDL. To use different QC thresholds and/or Smoothing and Joining settings, see "Setting QC parameters in the ChAS browser" on page 134.

# Changing segment filters (optional)

Click on the Filter icon (or click Files → Segments Filters).
 The Segments Filters window opens. (Figure 513)



- 2. Update the appropriate filters using the provided check boxes, text boxes and sliders.
- 3. Click **X** to save your changes and close the window.

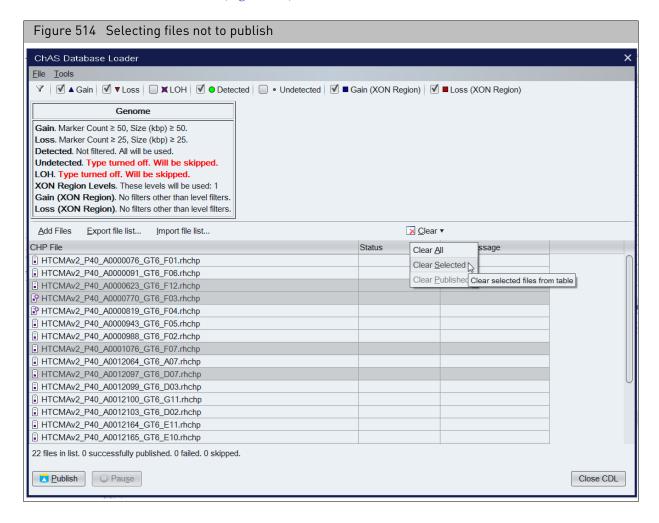
# Managing data types (optional)

- 1. Use the Segments Filters window (Figure 513) to click the check box of the data type(s) you want to publish.
- 2. Click **X** to save your changes and close the window
- 3. Review the Genome dialog window (Figure 512) again to make sure your data types to be published are displayed.



### Publishing your files

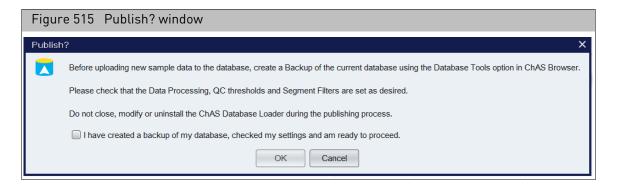
 Check your table before publishing, as all displayed files are published. Note: If there are specific files you do not want published, single click, Ctrl click, or Shift Click to highlight them, click the Clear drop-down menu, then click Clear Selected. (Figure 514)



**IMPORTANT!** You must have ChAS DB Manager or Admin privileges before you can publish. For information on setting up ChAS DB role assignments, see "Administration" on page 454.

2. Click Publish

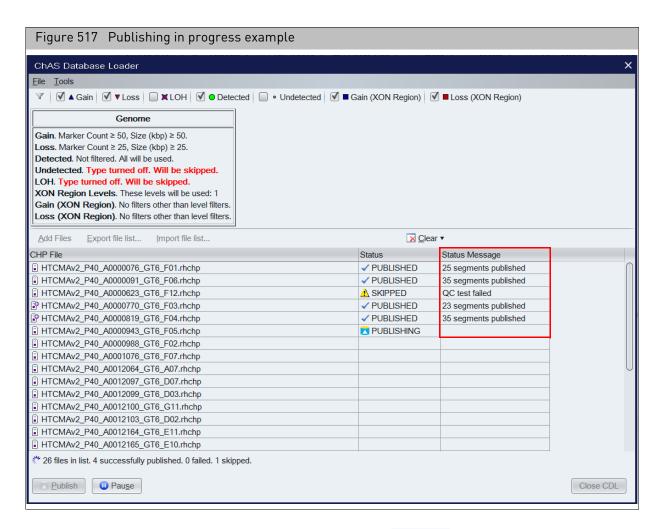
A **Publish?** window appears. (Figure 515)



- 3. Acknowledge the message, click to check its check box, then click **OK**.
- 4. An **Overwrite?** message may appear. (Figure 516) Click the appropriate button to continue.



The publishing process begins. (Figure 517)



To pause the publishing process, click Pause. While in pause mode, you can add more files to the table, as described in "Adding files to CDL" on page 460.

After the publishing process is complete, each **Status** column is marked with a result icon.

- = The file was successfully published to the ChAS database.
- The file was skipped over and not published, because it was already found in the database or it did not meet the assigned QC thresholds.
- X = The file failed and was not published.

**Note:** Refer to the table's Status Message column (Figure 517) for details regarding a skipped or failed file.

## **Clearing Table Data**

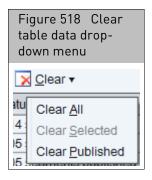
Clear Published

After publishing, click  $Clear \rightarrow Clear Published$  to remove all files that successfully published.

After clicking **Clear Published**, files with a Skipped or Failed status remain in the table. Click **Export properties...** to export these files for further investigation.

Clear All

Click  $Clear \rightarrow Clear All$  to remove all files from the table.



## Closing CDL

1. Click Close CDL or click File  $\rightarrow$  Close CDL.



## **Analysis parameters**

## Analysis parameters for single sample analysis

 For information on CytoScan algorithm parameters and their values, go to: http://media.affymetrix.com/support/developer/powertools/changelog/apt-copynumber-cyto.html

#### 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

• For information on CytoScan algorithm parameters and their values, go to: http://media.affymetrix.com/support/developer/powertools/changelog/apt-copynumber-cyto.html



## AED file format

Affymetrix Extensible Data (AED) files contain data that annotate positions on a genome. AED allows custom, typed fields, can be edited in the ChAS Browser's AED Editor feature ("Viewing and batch editing AED file contents" on page 307), and supports internationalization.

This appendix covers the formatting and use of Affymetrix Extensible Data (AED) files with ChAS.

- "AED file description"
- "Property name elements" on page 474
- "Compatibility" on page 478
- "References" on page 479

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

AED files can be produced and edited in:

- ChAS (Recommended)
- Text-editing software (Not Recommended)
- Spreadsheet software such as Microsoft Excel (Not Recommended)

## AED file description

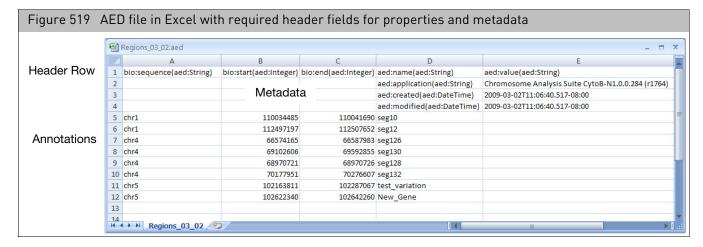
An AED file contains a list of annotations, descriptions of features on a biological *sequence* such as a chromosome. This description is comprised of several *properties*—either properties defined by this specification, such as the annotation start and stop positions; or properties defined by users 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 and has the following components (Figure 519):



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.)

- "Header row": Names the properties that can be used in the annotations
- "Metadata records" on page 473 Optional: Provides information about the AED file itself and the group of annotations it contains.
- "Annotations Rows" on page 474: The annotation row displays values for the properties listed in the header rows (for each feature that is annotated).

#### 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
  - A namespacePrefix is optional. It 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).

#### propertyIdentifier

– The 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

- The 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 must use this property if you are including metadata information in the file.

**IMPORTANT!** ChAS verifies 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 (Figure 520). 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.

igure o	20 Metadat	a entries			
io:seque	nce(aed:String)	bio:start(aed:Integer)	bio:end(aed:Integer)	aed:name(aed:String)	aed:value(aed:String)
				aed:application(aed:String)	Chromosome Analysis Suite CytoB-N1.0.0.284 (r1764
				aed:created(aed:DateTime)	2009-03-02T11:06:40.517-08:00
				aed:modified(aed:DateTime)	2009-03-02T11:06:40.517-08:00
Blanks for bio:sequence and other properties		Metadata Property names	Metadata Property values		

All other fields in a metadata record should be blank.

#### **Annotations Rows**

The rows below the Metadata properties are the annotations. Each row is a tabdelimited 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"
- "Properties" on page 475
- "Types" on page 477

#### 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:

Table 21 Namespaces

Namespace Prefix	Description	Examples
-	Custom and experimental properties not yet established as standard.	foo bar myProperty
aed	AED-specific properties and types.	aed:Integer
bio	Descriptions of biological entities.	bio:sequence bio:Strand
style	Information related to information representation, visually or otherwise.	style:color

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/, then 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.

#### **AFFX** 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.

Table 22 AFFX Properties

Name	Туре	Description
Affx:ucscGenomeVersion (aed:String)	aed:String	(Metadata) The genome assembly version using UCSC names, for example "hg19".

#### **AED** properties

These properties are parts of the AED namespace.

Table 23 AED Properties

Name	Туре	Description
aed:application	aed:String	(Metadata) 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	(Metadata) 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	(Metadata) The point in time the data was modified; this is not necessarily the time the file was modified.
aed:name	aed:String	(Required.) 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	(Required only if metadata records are present.) 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	(Metadata) 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**

Table 24 Biology Properties

Name	Туре	Description
bio:assembly	aed:URI	(Metadata) A URI indicating the genome assembly used. Currently the DAS GlobalSeqIDs 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.
bio:end	aed:Integer	(Required) 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	(Required) 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	(Required) 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.

Table 25 Style Properties

Name	Туре	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**.

Table 26 Types

Туре	Description	Lexical Form	Examples
aed:String	A sequence of Unicode characters.	character*	Abc a "fun" test!
aed:Integer	The positive whole numbers, the negative whole numbers, and zero.	[-] digit+	• 123 • 0 • -5000
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.	[-] digit+. digit+ [e[+ -] digit+] The lexical form "NaN" is explicitly prohibited.	• -123.0 • 0.0 • 123.45 • 1.2e+5
aed:Boolean	A binary true/false value.	"false"   "FALSE"   "true"   "TRUE"	• false • FALSE • TRUE
aed:URI	A Uniform Resource Identifier (URI).	See RFC 3986.	<ul> <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 ISO 8601 with required milliseconds and time zone offset.	YYYY-MM- DDThh:mm:ss.s+ (+ -)hh:mm	• 2008-09-12T18:45:43.779-07:00
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( <i>digit</i> +, <i>digit</i> +, <i>digit</i> +)	• rgb(0, 0, 0) • rgb(200, 50, 128) • rgb(255, 255, 255)

Table 26 Types

Туре	Description	Lexical Form	Examples
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.	• + • -
aed:UUID	A Universally Unique Identifier (UUID) as specified by RFC 4122 (in canonical form, not as a URN).	XXXXXXXX-XXXX- XXXX-XXXX- XXXXXXXXXXXX	• f81d4fae-7dec-11d0-a765- 00a0c91e6bf6

### 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 chromStart ≤ 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 (not recommended) 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

Though not recommended, 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

Though not recommended, 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 <a href="http://www.emeditor.com/">http://www.emeditor.com/</a> is a commercial text editor that has extremely good Unicode and BOM support, and is able to open up gigantic text files.

PSPad <a href="http://www.pspad.com/">http://www.pspad.com/">is a free text editor that has particularly extensive Unicode and BOM support and is available in many localizations.

UniPad <a href="http://www.unipad.org/">http://www.unipad.org/</a> 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

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



## 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 xxCHP 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 27 lists the namespace prefixes recognized within the ChAS user interface.

Table 27 ChAS namespace prefixes

Namespace URI	Prefix	Label
http://affymetrix.com/ontology/aed/	aed	General
http://affymetrix.com/ontology/aed/biology/	bio	Biology
http://affymetrix.com/ontology/aed/style/	style	Style
http://affymetrix.com/ontology/aed/default/	(none)	Custom
http://affymetrix.com/ontology/	affx	Affx
http://affymetrix.com/ontology/algorithm/	alg	Algorithm
http://affymetrix.com/ontology/algorithm/option/	algopt	Algorithm Option

Table 27 ChAS namespace prefixes

Namespace URI	Prefix	Label
http://affymetrix.com/ontology/algorithm/state/	algstate	Algorithm State
http://affymetrix.com/ontology/arr/	arr	ARR
http://affymetrix.com/ontology/chp/	chp	СНР
http://affymetrix.com/ontology/netaffx/	netaffx	NetAffx
http://affymetrix.com/ontology/genome.ucsc.edu/bed/	bed	BED
http://affymetrix.com/ontology/projects.tcag.ca/variation/	dgv	DGV
http://affymetrix.com/ontology/www.genome.ucsc.edu/fishClones/	fish	FISH
http://affymetrix.com/ontology/www.ncbi.nlm.nih.gov/omim/	omim	ОМІМ
http://affymetrix.com/ontology/www.pubmed.gov/	pubmed	PubMed
http://affymetrix.com/ontology/www.ncbi.nlm.nih.gov/RefSeq/	refseq	RefSeq
http://affymetrix.com/ontology/www.genome.ucsc.edu/genomicSuperDups/	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 as follows:

- 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 of the types listed above may not be available in some contexts.

## Automatic conversion of xxCHP headers to properties

The file format used by xxCHP files uses a separate approach for describing properties and value types. However, there exist equivalent AED types for the types used by xxCHP files. Moreover, the header names used in xxCHP files are logically grouped into categories by certain prefixes. ChAS uses these groupings to place xxCHP headers into AED namespaces when loading xxCHP 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, as shown in Table 28.

Table 28 CHP header name conversion to AED namespace

CHP Header Name Prefix	AED Namespace
affymetrix-algorithm-	http://affymetrix.com/ontology/algorithm/
affymetrix-algorithm-param-	http://affymetrix.com/ontology/algorithm/
affymetrix-algorithm-param-option-	http://affymetrix.com/ontology/algorithm/option/
affymetrix-algorithm-param-state-	http://affymetrix.com/ontology/algorithm/state/
affymetrix-chipsummary-	http://affymetrix.com/ontology/chp/summary/
(all others)	http://affymetrix.com/ontology/chp/

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 is 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, as shown in Table 29.

Table 29 Converted Properties

CHP Header Name	Property
affymetrix-array- type	http://affymetrix.com/ontology/arrayType
affymetrix- chipsummary-snp- qc	http://affymetrix.com/ontology/chp/summary/snpQC
affymetrix- chipsummary-MAPD	http://affymetrix.com/ontology/chp/summary/mapd

#### **Derived properties**

The following properties are each assigned to a file property derived from one or more header parameters, xxCHP 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)

- 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.
- ${\bf 3.}\ \ \, {\bf The\ affymetrix-algorithm-param-reference-mdl file\ header\ parameter.}$



# ChAS browser NetAffx Genomic Annotations

## Homo Sapiens database files

#### NetAffx Genomic Annotation files

#### Source of content

- NetAffx Genomic Annotation files are used by the ChAS Browser to display recent snapshots of genomic annotations downloaded from public databases.
- The UCSC Genome Browser is the source of the data that populates the following Browser tracks: Genes, Ensembl, Segmental Duplications, sno/ miRNA, and Cytobands. UCSC was also the source of OMIM data for ChAS Browser file versions NA31-NA33.1 and 32.1. NA33.2 and NA36 no longer contain annotation information for BACs and FISH Clones.
- The Database of Genomic Variants is the source of the data displayed in the DGV track.
- The ClinGen Resource is the source of the data displayed in the Triplosensitivity, Haploinsufficiency, and Recurrent CNV tracks.
- The OMIM database (with curation and processing assistance from UCSC) is the source of the data displayed in the OMIM Genes, OMIM Phenotype Loci, and OMIM Region Phenotype Loci tracks.

IMPORTANT! Starting with ChAS v3.3, the naming convention of the NetAffx Genomic Annotation database file will change. This file will now include a date (as opposed to a specific NetAffx build number) and will be updated more regularly than its current schedule of being updated with each software release. The new naming convention is as follows:

NetAffxGenomicAnnotations.Homo\_sapiens.hg38.naYYYYMMDD.db

For optimal use of the Segment Prioritization methods described in Chapter 17, "Prioritizing segments" on page 373, you must download the NetAffx Genomic Annotation files released with ChAS 4.3 or use more current ones when available.

**IMPORTANT!** It is VERY highly recommended to confirm findings obtained using the ChAS Browser's NetAffx Genomic Annotations file contents by linking out to external databases using the ChAS software coordinates for the most current annotation information. See "Linking to external websites" on page 215.



## Genomic position coordinates

There are multiple conventions and file formats to describe locations in chromosomal DNA sequences. This appendix describes a few issues that 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 "hg18", "hg19" and "hg38". The identical genome assemblies are known as "NCBI36", "GRCh37" 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 .

Consider the (randomly-chosen) SNP marker"S-3SRJC" from the CytoScan HD array. This marker is designed to correspond to the SNP with ID "rs4376202" in the dbSNP database. The NCBI website reports the position as chr4:1822637 on GRCh38. On the NetAffx website, the identical coordinate is also given for this SNP. 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 for this particular example.

For copy number markers which are not based on SNP positions, we continue to use a 1-based index position. 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 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

BED and AED file formats are used 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 Thermo Fisher Scientific 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 chromEnd = chromStart + 1. The UCSC browser does allow chromStart to be equal to 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.

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

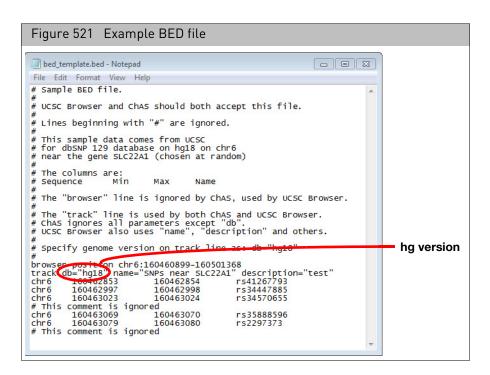
If there were a segment starting at markerA and ending at markerC, we would need to represent it in a BED or AED file as:

– Chr3 999 3000 segment\_1



## Editing BED files

A BED file is essentially a tab-delimited text file, as shown in Figure 521.



Using a text editor such as MS WordPad and MS NotePad (not a spreadsheet application like Microsoft Excel) to edit BED files is recommended and 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, which would cause the file to be invalid and unusable with ChAS or other applications.

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. You may use a text editor, but be certain to separate the columns with TAB characters and do not use non-ASCII characters. The BED format was not designed with such characters in mind; therefore, problems may occur when you try to share these files with others. ChAS will reject BED files containing non-ASCII characters, and will never export non-ASCII characters into a BED file.

- 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



# CytoScan 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.

**Note:** For CytoScan HTCMA algorithm and QC, see the RHAS User Guide.

Feature identification and signal extraction

GeneChip Cartridge Microarrays are scanned on the GeneChip Scanner and processed by the AGCC/GCC scanner software package. AGCC/GCC 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.

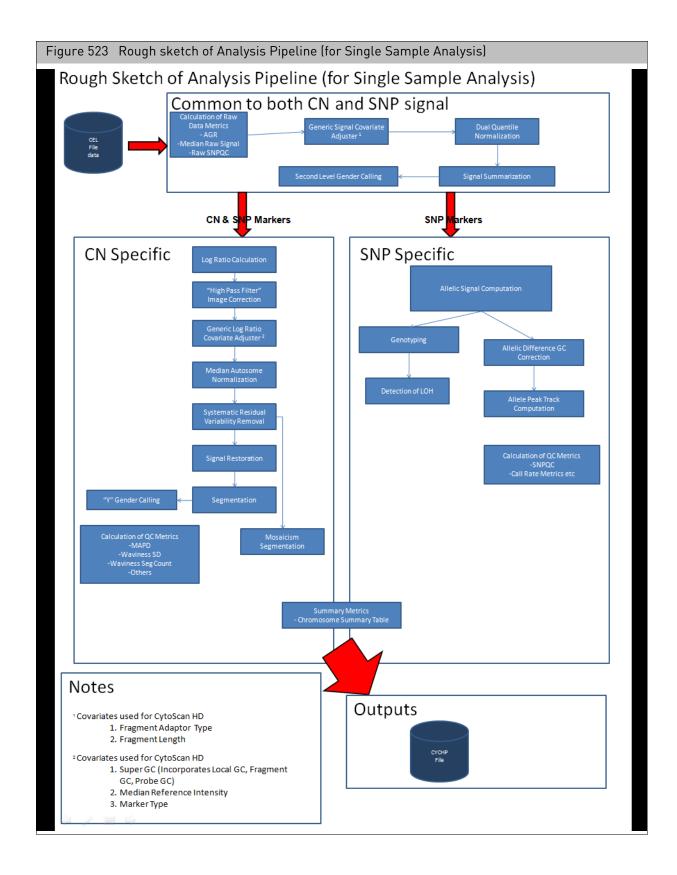
Single sample CytoScan workflow Beginning with the raw signal data in the CEL file, the Single Sample CytoScan Workflow implements a series of steps that perform probe set summarization, normalization, removal of variation caused by known properties and residual variation, and completing with calling genotypes, copy number segments and LOH segments.

A brief overview of each step performed by the CytoScan Workflow is shown in Figure 523 on page 493. In addition, a rough sketch of Analysis Pipeline (for Single Sample Analysis) is demonstrated in Figure 523 on page 493.

Figure 522 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 Fragment Adapter and Length Covariate Corrections **Dual Quantile Normalization SNP Probes** Log2 Ratio to Reference High-Pass Filter Image Correction Signal Summarization GC, Ref Intensity and Marker Type Alleleic Signal Computation **Covariate Corrections** Median Autosome Normalization Genotyping Systematic Residual Variability Removal Allelic Difference GC Correction Signal Restoration Allele Peak Track Computation Segmentation Detection of LOH

Visualization

Smoothing & Joining



## 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 Log2(sample<sub>m</sub>) – Log2(reference<sub>m</sub>), 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 arrays, 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.

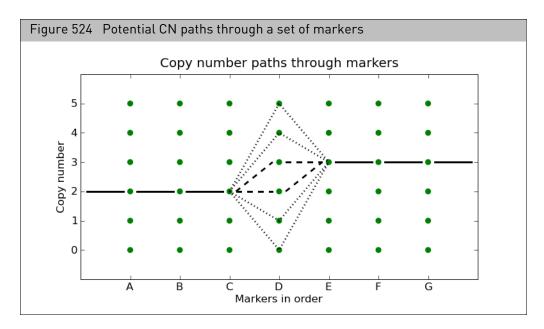
The weighted Log2 Ratios are centered on copy number (CN) = 2. In theory, when Log2 Ratio = 0 then CN = 2, when Log2 Ratio = -1 the CN = 1, etc. In truth, microarrays, or any hybridization-based technology, exhibit Log2 Ratio compression due to many factors, so the Log2 Ratios never exhibit the amplitude expected by the math. The following table shows theoretical and actual Log2 Ratios for different Copy Number States.

Table 30 Copy number states and Log2 Ratios

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 (Figure 524), there are many potential paths through the possible CN states of 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 probably 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 done 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 ≤50 markers and ≤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.

## 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 3MB or more in size (for CytoScan HD). Copy number change events less than this size may be detected; however, sensitivity and specificity will be reduced.

#### Limitations in mosaicism segmentations

- The algorithm is designed to detect only mosaicism between approximately 15%-70% mosaicism for copy numbers between 1 and 3 for regions on the order of 3MB in size or larger for autosomes and chrX.
- The algorithm is designed to detect only mosaicism between approximately 15%-70% mosaicism for copy numbers between 0 and 2 for regions on the order of 3MB in size or larger for chrY.
- The algorithm will not call events below 1MB in size irrespective of number of markers.
- The algorithm will not call events below 10% mosaicism irrespective of genomic size or number of markers.
- The algorithm does not use allelic difference or BAF data for segmentation purposes. It is recommended that allelic difference and/or BAF data tracks be examined in conjunction with the reported mosaicism segmentation.

### SNP marker workflow

## Signal summarization

CytoScan arrays contain six probes for each SNP probe set, three 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 arrays is accomplished using the BRLMM-P algorithm described in the White Paper: BRLMM-P: A Genotype Calling Method for the SNP Array 5.0 (2007).

## Allelic difference GC correction

Systematic changes in Allelic Differences can be related to differences in GC content. For instance, on a given sample Allelic Differences representing AA and BB genotype markers might get progressively closer or further from each other as the GC content changes. It is assumed that such changes represent unwanted variability. The Allelic Difference GC correction determines differences in the structure of the allelic differences associated with GC and then removes these differences. For CytoScan HD the super GC covariate is used.

#### **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 While Paper: The Loss of Heterozygosity (LOH) Algorithm in Genotyping Console 2.0.

## Array data QC metrics

This section provides a high level overview of the key QC metrics used with the CytoScan arrays.

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 Log2 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 (for CytoScan 750K and HD, MAPD > 0.29 for CytoScan Optima) 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 (for CytoScan 750K and HD, SNP QC < 8.5 for CytoScan Optima) is of poorer quality than is required to meet genotyping QC standards.

## ndSNPQC (SNP Quality Control of Normal Diploid Markers)

The metric, SNPQC is a measure of how well genotype alleles are resolved in the microarray data. ndSNPQC is the same metric but only applied to normal diploid markers (that is those that have been determined to have Copy Number =2 in the sample). Larger ndSNPQC values are better.

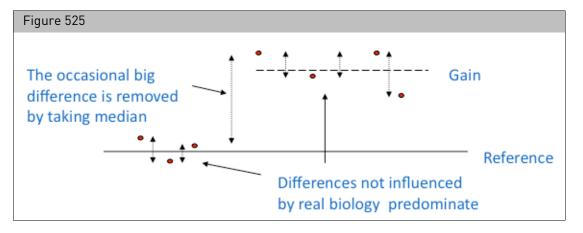
## ndWavinessSD (Normal Diploid Waviness Standard Deviation)

ndWavinessSD is a global measure of variation of microarray probes that is insensitive to short-range variation and focuses on long-range variation. ndWavinessSD is computed on normal diploid markers.

## 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 (Figure 525).



Formally, if xi: is the Log2 Ratio for marker i:

MAPD = median( $|x_{i-1} - x_i|$ , with *i* 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 MAPD/0.96 is equal to SD and MAPD\*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.

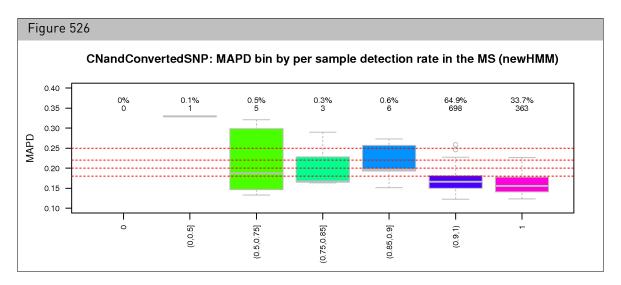
Variability in Log2 Ratios in a microarray arises from two distinct sources:

- 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. (Figure 526)



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 Log2 Ratios across the whole genome and subtracting out the short-range variation, specifically, for autosomal probes:

Define:

X<sub>i</sub> as the Log2 Ratios of autosomal probes

And Z<sub>i</sub> 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  $(Z_i)$ :

Waviness-SD =  $sqrt(Var(X_i)-Var(Z_i)/2)$ 

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

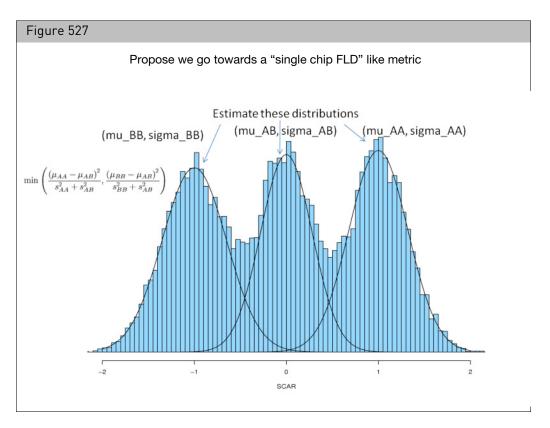
multiple-chromosome constitutive trisomies, where a considerable fraction of the genome is duplicated.

For most samples, a Waviness-SD value below 0.12 for CytoScan arrays indicates that the long-range variation is within levels that can be accommodated by the CytoScan 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.

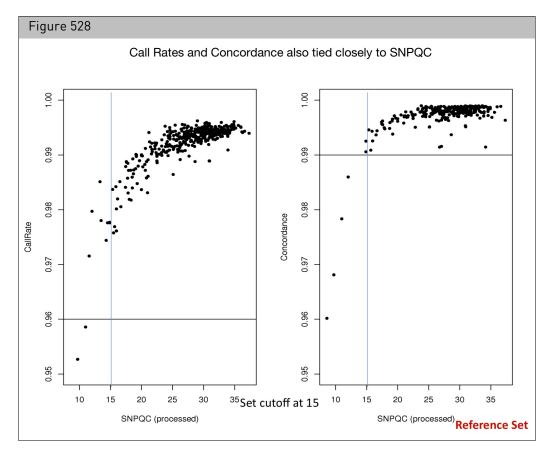
Waviness-SD can be a good indicator of process drift since it measures long-range variation relative to the CytoScan HD or CytoScan 750K reference profile. A general rise of Waviness-SD for all samples coming from your laboratory may be an indication of a change of protocol, technique or reagents.

## **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 (Figure 527).



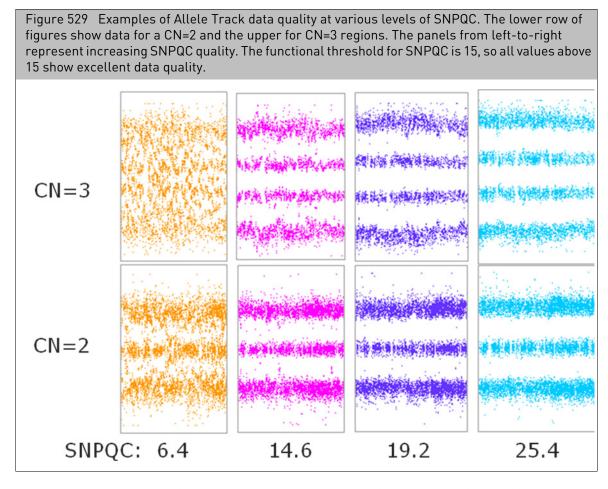
SNPQC correlates well with genotype performance, as measured by Call Rate and Concordance to published HapMap genotypes. To establish this relationship, we scored 380 microarrays from the Reference Set by calculating SNPQC, Call Rate and Concordance. The following graphs show the relationships between SNPQC and the other two metrics (Figure 528).



The left panel shows that when SNPQC > 15, Call Rate is above 98%. The right panel shows that when 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 SNPQC > 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 529).



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.

**Note:** For detailed information on algorithms and QC metrics for the OncoScan array, refer to the **OncoScan Console User Manual** (P/N 703195).

## TuScan algorithm

The TuScan algorithm uses B-allele frequencies (BAFs) and log2 ratios to estimate the ploidy and percentage of aberrant cells in the sample (%AC) which in turn are used to calculate copy number calls (CN). The BAFs and log2 ratios contribute equally to CN determination. TuScan first uses the BAFs and log2 ratio data to identify segments of equal CN. Next TuScan uses the BAFs, log2ratios and segment data to find the combination of %AC and ploidy that best fits the data. When TuScan can successfully determine %AC, the algorithm assigns each aberrant segment an integer copy number representing the copy number in the tumor portion of the sample. This is possible because CN is well approximated by an integer when the tumor is nearly homogeneous. If the tumor is highly heterogeneous (i.e., lacks a dominant clone), or contains a large amount of "normal" cells "AC cannot be determined. In other words, if the percentage of aberrant cells contributing to the various aberrations in the sample varies across all aberrations, %AC and ploidy cannot be determined. When %AC cannot be determined, the segmentation algorithm will still identify segments of equal CN, but the CN in just the aberrant cells cannot be determined. In this case, TuScan bins the copy numbers and returns fractional CN values in 1/3 increments (e.g., 2, 2.33, 2.66, 3 etc.). This fractional copy number is derived from the normal contamination as well as the heterogeneous population of tumor cells; therefore, the fractional CN calls represent the average CN observed for that segment. Users should look at the value of %AC to determine whether the CN value represents the CN in the tumor (%AC= number) or the average CN in the sample (%AC=NA). Tumor heterogeneity also affects the interpretation of the CN number calls when %AC cannot be determined. For example, a TuScan call of 2.33 can result from 40% of the aberrant cells having 3 copies, 10% of aberrant cells having 5 copies, or a more complex heterogeneous mixture of copy numbers. Since nearly every tumor sample will have some amount of normal contamination combined with tumor heterogeneity it is not possible to predict how often TuScan will be able to determine the %AC, it will vary depending on the sample.

## Manual re-centering algorithm (OncoScan)

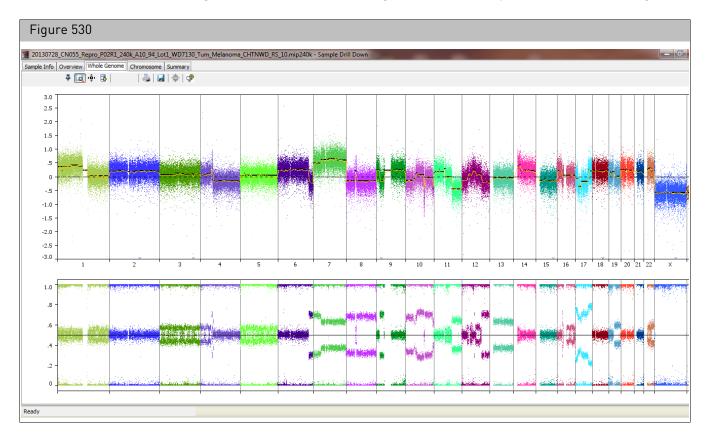
TuScan identifies normal diploid markers in a sample of interest, determines the copy number for these markers (2, 4 or 6) and ensures that markers with CN=2 have a log ratio of 0. This is referred to as "centering" the sample.

When no or an insufficient number of normal diploid markers are found, the automatic recentering does not occur. In addition, occasionally the automatic recentering misses the true CN =2 markers and does not correctly center the sample. In these cases, it is advised to center the sample manually to get correct CN calls. Manual recentering is now available through the CHAS software and the recentered sample is re-run through TuScan (described above) to provide integer copy number.

The new RC.OSCHP files can be viewed in ChAS or the BioDiscovery software, Nexus.

To manually recenter samples, an offset (median log2 ratio) is provided that tells the algorithm how much a sample should be pushed up (positive value) or pushed down (negative value) so that this region resides at the log 2 ratio = 0, indicative of normal diploid.

In the example below (Figure 530) the sample should be centered at chromosome 4q. The median log ratio on 4q is -0.17, therefore the manual recentering adjustment would be given this offset value, resulting in an increment adjustment of 0.17 for all log ratios.

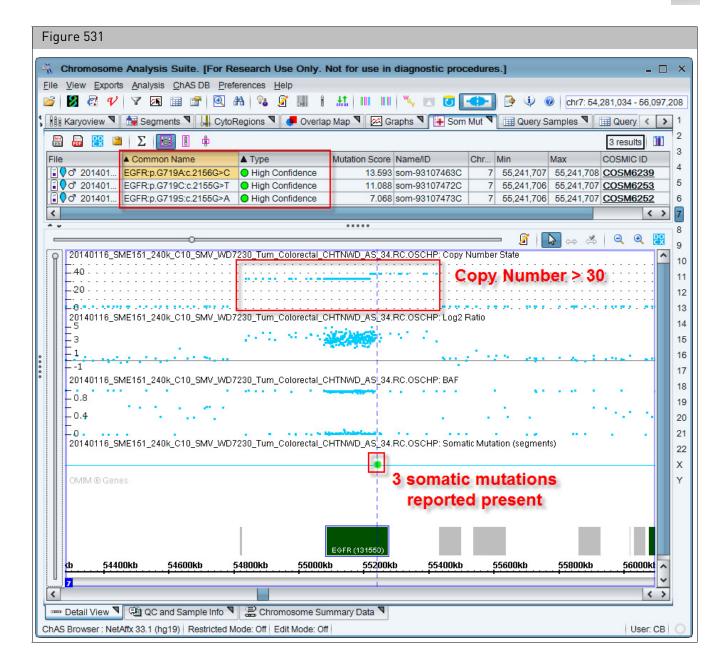


## Copy number effect on somatic mutations

Somatic mutation probesets in the OncoScan FFPE Assay are designed to selectively respond to the presence of mutation sequences. However, large copy number amplifications spanning the somatic mutation targets can sometimes lead to falsely reporting the presence of mutations in amplified regions.

If the copy number state is greater than ~15, you may observe false positive somatic mutation calls. The only region for which we have observed this problem is the EGFR gene, which is prone to very high copy number in certain cancer types.

In the example below (Figure 531), the predicted copy number state for the EGFR gene is greater than 30, which affects the somatic mutation score. Another side effect shown in the example below is that three mutations are called in the high Copy Number region, a contradictory event for at least two of these mutations.



## CytoScan XON region calling algorithm

For CytoScan XON arrays a region focused calling algorithm was developed to distinguish between neutral, gain and loss copy number. Regions are defined by genomically contiguous stretches of closely spaced markers. Each region is assessed independently.

A three state conditional random field (CRF) model is used to segment each region. The CRF model uses the log2 ratio measurements corresponding to that marker and four markers on either side, plus the state of the previous marker, to assess whether a given marker is in one of the three states. Before using the CRF model the log2 ratio is independently adjusted for every marker to account for expected differences in responsiveness to copy number.

For resistance to outliers, the t-distribution is used for computation of emission probabilities. The mean parameter for the emission distributions are fixed for each state across all markers, but the standard deviation parameter includes a component that differs for each marker, based on expected marker variability. Adjacent regions with the same copy number determination are not joined into larger segments so large copy number aberration events may be represented by multiple region based calls.



# Recommended CytoScan XON array workflows

**IMPORTANT!** The Segment Prioritization feature outlined in Chapter 17, "Prioritizing segments" on page 373 can be used in conjunction with the recommended steps below to quickly prioritize XON regions.

## Whole exome analysis

The markers on the CytoScan XON array have been categorized into these four levels (based on the annotation in the region of the genome).

- Level 1: Includes genes with the highest level of evidence: developmental delay, epilepsy, ASD, XLID, Metabolic disorders, hereditary cancer OMIM Morbid genes.
- Level 2: ClinVar genes not covered in Level 1.
- Level 3: Other OMIM genes not identified as Level 1.
- Level 4: Other Ref Seq, UCSC, Ensembl genes, LOVD.

Recommended workflow for analyzing the whole exome

#### To load XNCHP file(s):

- 1. Click **File**  $\rightarrow$  **Open**.
- 2. Select the XON-Level 1 Named Setting.

The XON Region Segment Calls in genomic regions assigned as Level 1 are shown.

3. Turn off any filters that are based on markers or size.

The XON gain/loss segments that are contained in/or overlap with Level 1 regions will be exposed on the XON Region Segment Track. The data in the Level 1 regions (Log2 Ratio, Weighted Log2 Ratio, Smooth Signal, Allele Difference, Ballele Frequency) will show in color (based on the sample color assignments in the User Configuration). The data in the remaining Level 2-4 regions will remain gray and no XON segments will be revealed.



- 4. Add the XON Region Level and Summarized Log 2 Ratio columns to the segment table, then use the Segments Table to review the XON Region segment calls.
- 5. Optional: Annotate the XON Region segments as you normally would using the Call and Interpretation columns, as described in "Adding annotations at the sample (xxCHP) file level" on page 251.
- 6. Optional: Exon Region segment calls from other Levels can be displayed by selecting the check box(es) in the Filters Tab.



7. Optional: Publish the XNCHP file to the ChAS DB, as described in "Publishing data to the database" on page 404.

#### Targeted analysis

When you have a specific set of genes that you want to search for XON Region segment calls, a targeted analysis can be performed using CytoRegions for Targeted XON analysis. Details on this feature can be found in "Assigning a CytoRegion for targeted XON analysis" on page 278.

- Create an AED file from your list of genes. See "Creating an AED File from a gene list" on page 278. If you already have an AED/BED file containing your genes of interest, go to step 2.
- 2. Click **File**  $\rightarrow$  **Open** to load the AED/BED file.
- 3. Locate the file, then click **Open**.
- 4. Right-click on the loaded file, then select Set File as CytoRegion for Targeted XON Analysis.

This assigns the file as a CytoRegions File, as well as automatically sets the appropriate filters for Targeted XON analysis. The filters hide all segments in the Genome region and colors all the data gray and turns on all levels in CytoRegions to expose any XON Region segment call and colors the data the assigned sample color.

**Note:** Only XON Regions calls within the CytoRegions are displayed in the views and tables.

- 5. Optional: Annotate the XON segments as you normally would, as described in "Adding annotations at the sample (xxCHP) file level" on page 251.
- 6. Optional: Publish the XNCHP file to the ChAS DB, as described in "Publishing data to the database" on page 404.

