CytoSure™

ogt

CytoSure Interpret Software

User Guide

Oxford Gene Technology

Founded by Professor Sir Ed Southern, Oxford Gene Technology (OGT) supplies world-class genetics research solutions to leading clinical and academic research institutions.

- **CytoSure™** — Class-leading products and services offering the complete array solution for clinical genetics research.
- **Cytocell®** — High-quality fluorescence *in situ* hybridisation (FISH) probes for the detection of gene rearrangements related to inherited genetic disease and cancer.
- **SureSeq™** — Hybridisation-based NGS panels and library preparation reagents delivering flexibility and unparalleled coverage uniformity for clinical research applications.

For more information, visit www.ogt.com.

Contents

Introduction

CytoSure Interpret Software is a powerful and easy-to-use package for the analysis of aCGH data, offering an impressive combination of features that allow you the choice of standardised data analysis (using the Accelerate Workflow) or customised, user-defined data analysis. Additional functionality allows effortless identification of both copy number variation (CNV) and loss of heterozygosity (LOH) when used in conjunction with CytoSure arrays containing single nucleotide polymorphisms (SNPs). CytoSure Interpret Software is exclusively provided with Oxford Gene Technology's (OGT) extensive range of CytoSure aCGH arrays.

CytoSure Interpret Software delivers:

- Fast, accurate and simple analysis of aCGH data
- Comprehensive data annotation with direct links to external databases and online resources
- Robust relational database allowing sophisticated data querying and filtering
- Extensive customisation options to meet the needs of your laboratory
- Fully integrated, automatic analysis of array image files

Fast, accurate and simple analysis of aCGH data

CytoSure Interpret Software utilises an innovative "Accelerate Workflow" that provides automated data analysis based on predefined settings. This unique feature minimises the need for user intervention and maximises the consistency and speed of the data analysis. The analysis time is further reduced by using a batch processing facility, which allows an unlimited number of samples to be analysed simultaneously using the robust, optimised Circular Binary Segmentation (CBS) algorithm for identifying chromosomal abnormalities (Figure 1). The proprietary SNP calling algorithm and associated thresholds enable accurate calling of regions of LOH (Figure 2). In addition, the bulk load feature allows faster sample loading as multiple sets of sample information can be loaded at once as a text file.

Analysis can be performed using arrays designed to human genome build 18 (hg18) or 19 (hg19) — with easy, one-click switching between builds. The future-proof database structure allows further new builds to be easily imported and existing data to be re-mapped.

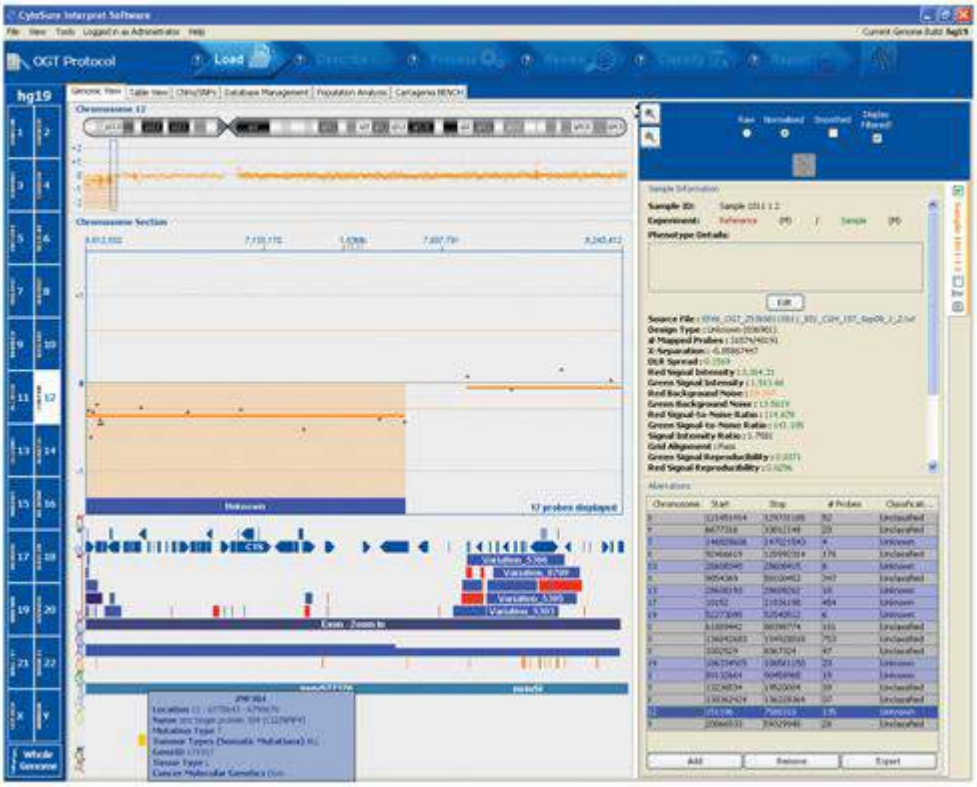


Figure 1: Automated aberration detection with CytoSure Interpret Software, showing clear detection of chromosomal abnormalities. The gain on chromosome 12 for this chronic lymphocytic leukaemia (CLL) sample contains the zinc finger protein gene ZP384, easily identified in the Cancer gene census genes track.

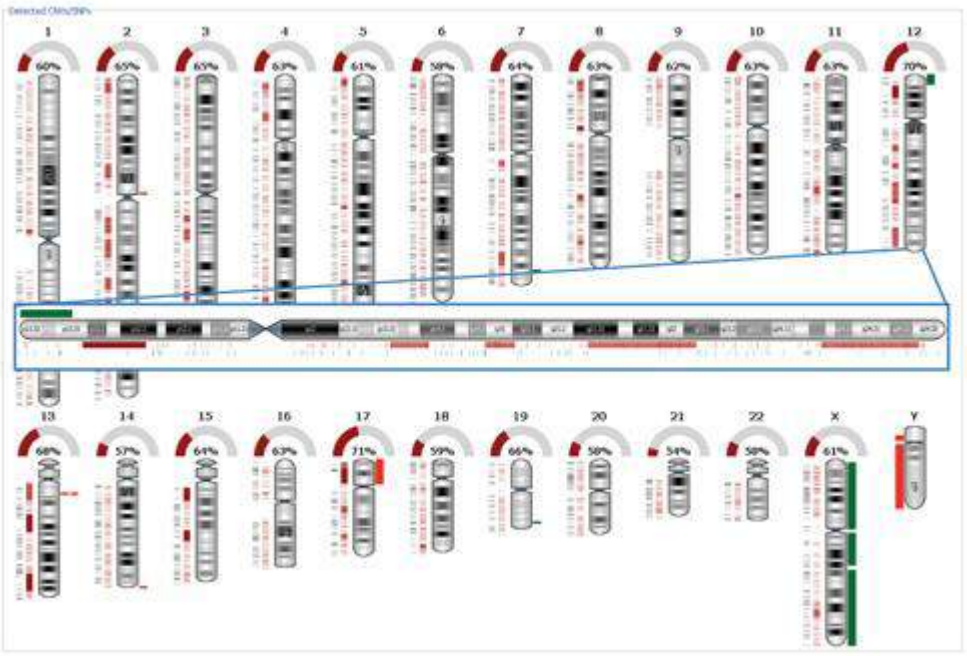


Figure 2: Automated SNP detection and LOH calling with CytoSure Interpret Software. The dark red rectangles indicate regions of LOH. The green and the bright red rectangles indicate amplifications and deletions respectively. This is the same CLL sample as displayed in Figure 1 and clearly illustrates the gain in the telomere region of the p arm and the region of LOH in p13.31-p12.3.

Comprehensive data annotation with direct links to external databases and online resources

CytoSure Interpret Software includes extensive annotation tracks covering syndromes, genes, exons, CNVs and segmental duplication — each of which link to publicly available databases such as ISCA, Decipher, Database of Genomic Variants and the Cancer Gene Census (Figure 3) providing results in context. Which tracks are displayed is fully customisable allowing only tracks of specific interest to be viewed (e.g. cancer-specific tracks) ensuring easy data interpretation. Each track can reference hg18 or hg19 information. Annotations within a track can be coloured allowing easy visualisation. It is also possible to customise what information from the tracks is saved in the report (e.g. how many common variants overlap with an aberration). In addition, CytoSure Interpret Software seamlessly integrates with Cartagenia's BENCH™ software, allowing users to securely transfer CNV calls for aCGH aberration data management and genotype-phenotype correlation, providing greater confidence in the calls made.

Figure 3: Fully customisable tracks simplify interpretation of aberrations.

Robust relational database allowing sophisticated data querying and filtering

The powerful relational database enables storage of sample data according to its relationship with other data. For example, samples can be segregated into different projects by sample type or by user, thereby simplifying analysis as only the data relevant to your work is visible. Annotation of individual samples is defined by the user; custom fields enable information applicable to your samples and experimental set-up to be stored within the database. The information contained within these custom fields can then be added to the final report. The database can be accessed simultaneously by multiple users.

Back-ups of the database are straightforward; simply choose a full, partial or mini back up, creating a compressed file which can be reinstated if necessary using the restore function. The installation of the relational database is customisable with a choice of database management systems designed to integrate with your current IT infrastructure.

The relational database allows flexible data searching utilising “Quick” or “Advanced” search functionality. Advanced search uses Boolean operators to query sample and experimental details as well as aberrations that are stored in the database. The search criteria can be saved allowing searches to be easily repeated.

CytoSure Interpret Software also makes it easy to examine inheritance patterns within related samples using a unique “Family Tree” viewer. Probands can be linked to parents and other family members with the facility to view three generations together (Figure 4).

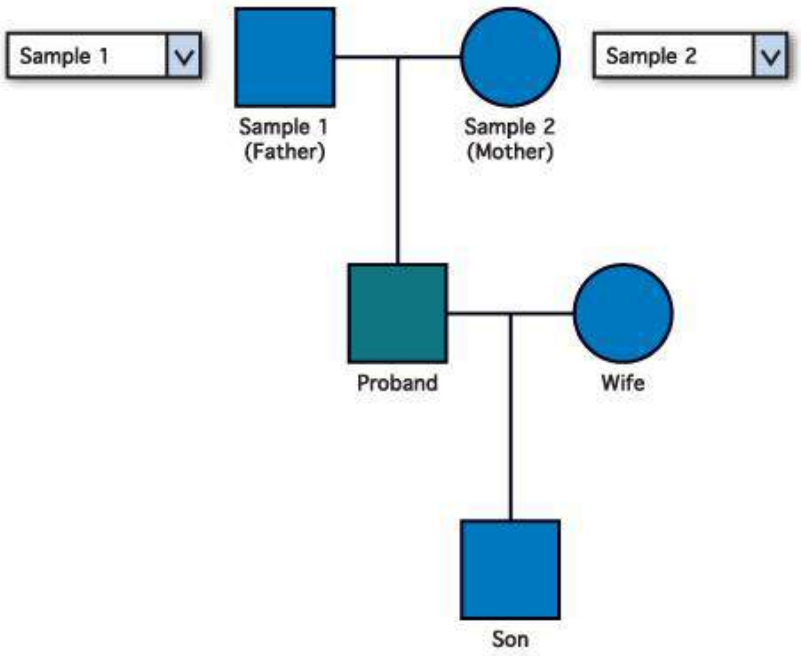


Figure 4: Pedigree information can be calculated using the novel Family Tree viewer.

Extensive customisation options to meet the needs of your laboratory

CytoSure Interpret Software gives you the flexibility to optimise data analysis settings for your laboratory's specific needs as well as the freedom to customise data reports to show only the information you need. In addition, the permission-based log-on structure enables greater flexibility with respect to management of user accounts. Within an organisation an Administrator can control the access and functionality available to users. This ensures consistency of analysis and reporting and provides an audit trail of any changes made within the software.

Fully integrated, automatic analysis of array image files

CytoSure Interpret Software can be used with the CytoSure Interpret Feature Extraction Module to analyse TIFF images from a variety of commonly available microarray scanners, including NimbleGen, Agilent, Innopsys and Axon. The module comes pre-loaded with template files enabling images to be feature extracted and seamlessly loaded into the Accelerate Workflow without the need for user intervention. The CytoSure Interpret Feature Extraction Module will also automatically search for TIFF images as they are generated, making it possible to start scanning, walk away and return to fully analysed data — a perfect fit to scanning at the end of the working day and interpreting results the following morning.

Alerts can be set to ensure that only good quality images are analysed, informing the user of any images that may need to be visually inspected. As well as streamlining the analysis workflow and reducing hands-on time loading and transferring data, the CytoSure Interpret Feature Extraction Module is agnostic regarding the source of TIFF images. It is also possible to scan a high-resolution microarray with a small feature size (i.e. 30 µm) at a low resolution (5 µm), widening the array choice for users with 5 µm scanners.

Constant innovation ensuring class-leading data interpretation

Constant innovation ensures that CytoSure Interpret Software remains the industry-leading aCGH analysis software. Our dedicated software development team are continually implementing enhanced features and capabilities in response to new product development and customer feedback. To complement our range of arrays containing both CNV and SNP probes, CytoSure Interpret Software can identify DNA regions exhibiting loss of heterozygosity (LOH), whole chromosome uniparental disomy (UPD) and segmental UPD. In addition, potential mosaicism can now be detected using our specialised threshold factors. Another unique feature is the facility to identify sample mix-up when used in conjunction with CytoSure Sample Tracking Spike-ins — providing confidence in results. This constant innovation ensures you can rely on CytoSure Interpret Software to provide fast and reliable delivery of meaningful results.

Administration and configuration

CytoSure Interpret Software contains extensive customisation options, including the flexibility to optimise data analysis settings for your laboratory's specific needs and the freedom to customise data reports to show only the information you need. Some of these settings are configuration options which control aspects of the software behind the scenes. Often these settings are set according to laboratory requirements and are standardised across the team to ensure all data is analysed in the same way.

Settings are associated with a Windows account on a specific computer (not a CytoSure Interpret Software user) and can be changed, loaded or shared by certain CytoSure Interpret Software users. Additionally, some users may have access to some of the settings (with others are greyed out), and are able to load Shared Settings, and others may not have access to any of the configuration settings. Whether these and other actions are available to a user is determined by their permissions — see User Account Administration for more details.

The following sections of the guide explain how to access all of these configuration options as well as how to share settings between users (Sharing Configuration Settings). All configuration options are accessed via **Tools -> Options...** (Figure 1).

See User Account Creation for more information on user levels (Admin, Super User, Read and Write and Read Only users).

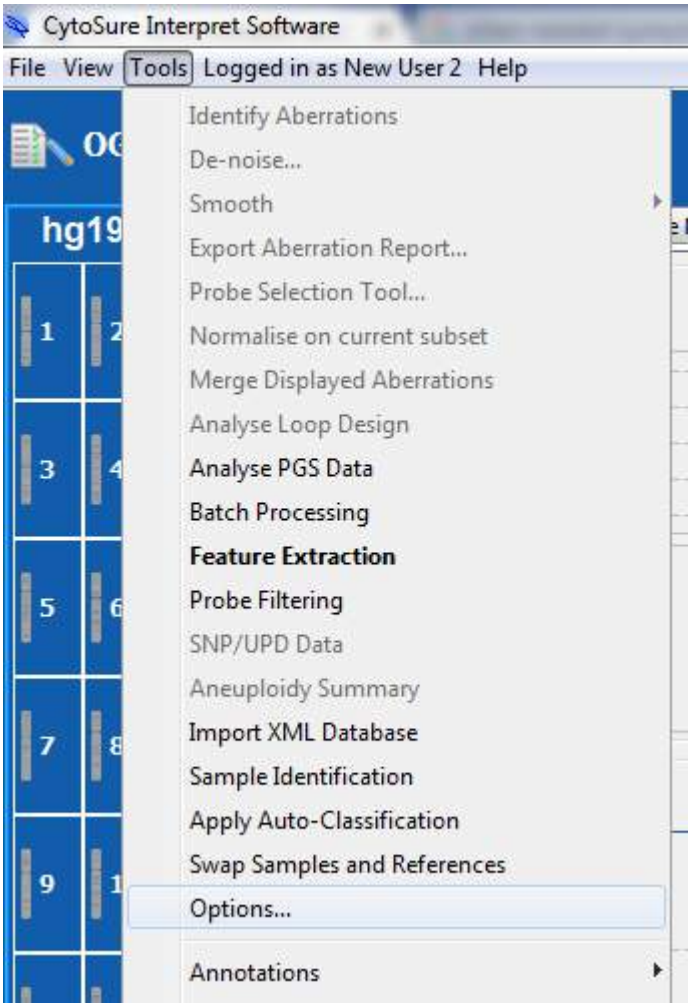


Figure 1: How to access the Options window through the Tools button in the toolbar.

Sharing configuration settings

It is possible to share between users a set of all configuration options. To do this, one user must share their settings (so they are saved to the database) and another user must load the shared settings to their own account. By default, Administrators and Super Users have permissions to share and to load options, Read and Write users have permission to load only and Read Only users do not have either of these permissions. If it is necessary to alter the configuration settings for a Read Only user's computer account, it would be necessary for another user type to log into CytoSure Interpret Software on the Read Only user's computer account and edit the settings.

> The configuration settings are saved in the Windows user home directory on the local machine, meaning that they are tied to that Windows account and not to the CytoSure

Interpret Software user (i.e. Switching CytoSure Interpret Software user will not affect the configuration settings).

In order to share, load and delete settings, follow these steps:

1. Choose **Tools -> Options... -> Database** to navigate to the database tab of the configuration options window (Figure 1 below).
2. To share the current settings, click **Share** in the **Shared Settings** section of the tab. A new row should then appear in the **Shared Settings** table with your user name and the date.
3. To load settings (usually from a different computer and/or logged in as a different Windows user), click on the relevant row in the **Shared Settings** table and click the **Load** button.
4. To delete a loaded settings file, click on the relevant row in the **Shared Settings** table and click the **Delete** button. The entry should disappear from the table.

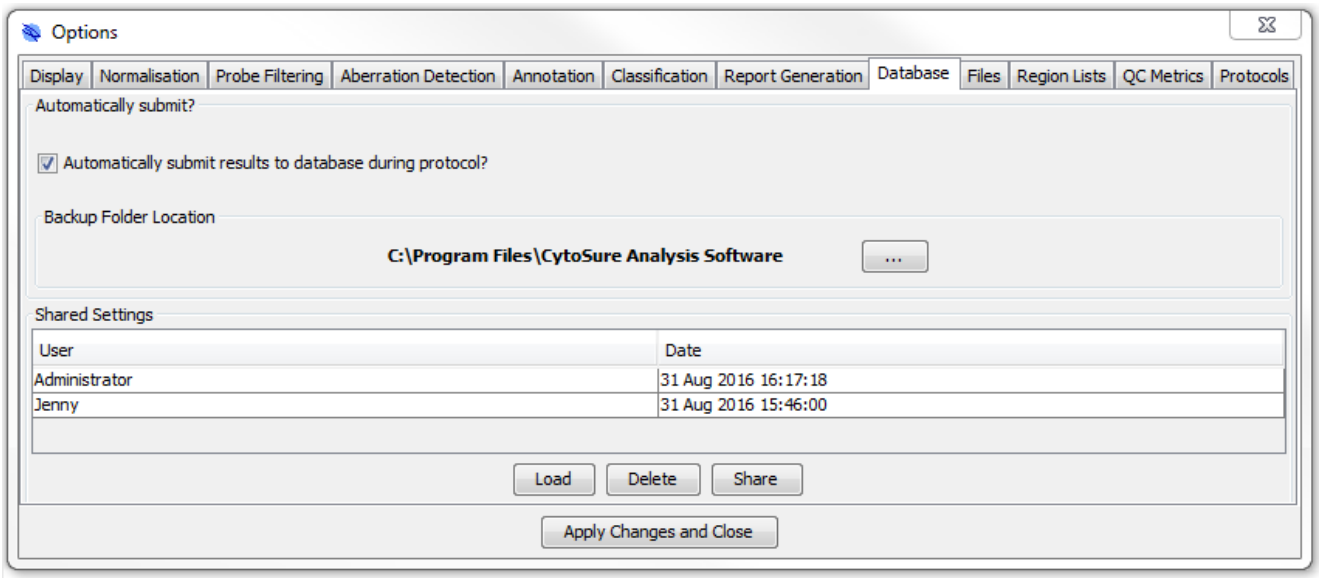


Figure 1: Database tab of the configuration options window.

User account administration

CytoSure Interpret Software uses a permission-based log-on structure. This enables flexibility with respect to management of user accounts, including the ability of an Administrator to control the access and functionality available to other users. This

functionality ensures consistency of analysis and reporting as well as providing an audit trail of certain actions carried out by users within the software.

The following sections describe the creation and management of user accounts and their associated permissions. All functionality is accessed through the **Logged in as [Username]** option in the toolbar (Figure 1 below).

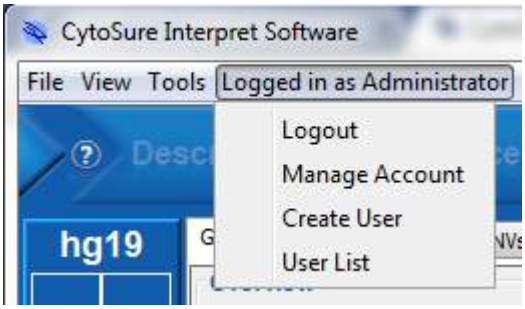


Figure 1: Logged in as [Username] toolbar options.

Creation

During installation, the administrator account is automatically created using the following details:

- User Name: *admin*
- Password: *cytosure*

> The admin account password should be changed the first time the software is used. See Management for help doing this.

From this account, it is possible to create additional user accounts. To do this, select **Logged in as Administrator** from the menu bar and choose **Create User**. A window will open (Figure 2) into which the new user details should be added. Details include: name, user name, user level, lockout time and account password. Once all details are added, press the **Create User** button at the bottom of the window to save the new user. The user should then be able to log-in with the details provided.

Depending on the user level of a new user, they may also be able to create new user accounts. The process for this is the same as for an administrator, however they are only permitted to create user accounts with user levels lower than their own (see User Levels).

> All user accounts must have a unique user name.

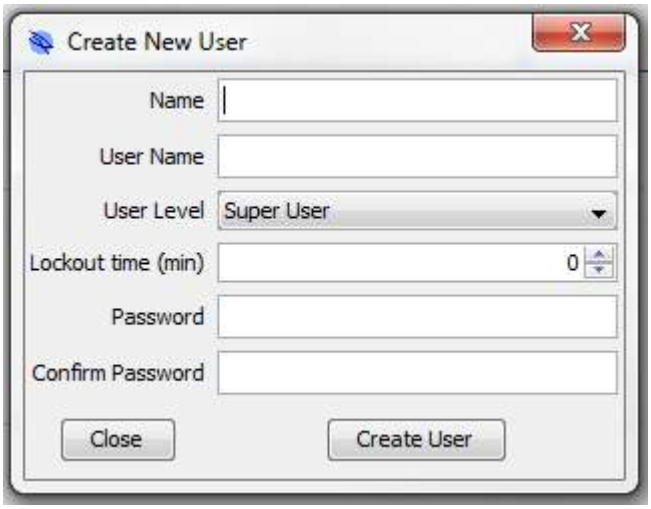


Figure 2: Create New User window.

User levels

As an administrator, there are three possible user levels to choose from when creating a new user. Each user level type is associated with different sets of basic permissions:

1. **Super User** — Users have all possible permissions except the ability to create super user accounts.
2. **Read and Write** — Users can edit the protocol settings and change the classifications of an aberration that has been submitted to the database. These users can also create Read Only user accounts (following the same process as an Administrator).
3. **Read Only** — Users are not able to make any changes to the protocol settings or the classification of aberrations. Users cannot create user accounts.

Table 1 lists the default permissions associated with each user level. These permissions can be edited after the creation of the user without altering the user level, as described in Management. Please note that permissions associated with creating other users are not listed as they cannot be edited.

Permission Type	Admin and Super Users	Read and Write Users	Read Only Users
Create Protocols	✓	×	×
Submit to Database	✓	✓	×
Re-Submit to Aberration Database	✓	✓	×
Modify Aberration Database Entries	✓	✓	×
Archive Protocols	✓	×	×
Add Annotations	✓	×	×
Add/Edit/Delete Categories and Fields	✓	×	×
Create/Edit File Parsers	✓	×	×
Load Data from Text File	✓	✓	×
Add/Edit/Delete Metric Threshold Sets	✓	×	×
Create Samples	✓	✓	×
Delete Samples	✓	×	×
Edit Sample Details	✓	✓	×
Create Experiments	✓	✓	×
Delete Experiments	✓	×	×
Edit Experiment Details	✓	✓	×
Change Default Experiment	✓	✓	×
Modify Analysis Protocol Options	✓	×	×
Modify General Options	✓	✓	×
Change Genome Build	✓	✓	×
Set Initial Classification	✓	✓	×
Set Final Classification	✓	✓	×
Edit Annotation Track Details	✓	×	×
Delete Annotation Tracks	✓	×	×
Edit Classification Types	✓	×	×
Add/Edit/Delete Annotation Links	✓	×	×
View Audit Trail	✓	×	×
Load Shared Settings	✓	✓	×
Share Settings	✓	×	×
Delete Shared Settings	✓	×	×
Change Database CNV Classifications	✓	×	×

Table 1: Default permissions for each user level.

Management

It is possible to change the following details of an existing user account:

- Name
- User name
- User level
- Password
- Specific permissions

Some details cannot be changed!

- Users can only alter any details for an account which is beneath them in the user hierarchy (i.e. created by them, created by a user created by them, or created by a user created by a user created by them, etc.).
- Users cannot change the user level or specific permissions of their own account.
- It is never possible to alter the user level or specific permissions of an administrator account.

When possible, changes are made through the **Manage User Account** window (Figure 3). If a user needs to manage their own details, this window should be accessed by selecting **Logged in as [user name] -> Manage Account**. If a user needs to manage the details of another user, they access this window by selecting **Logged in as [user name] -> User List**, choosing the relevant user from the list (Figure 4) and selecting **Manage User** from the bottom of the list window. If the user does not have permissions to manage the user chosen, an error message will be shown. If the user does not have permission to edit certain details only, the relevant fields will be greyed out in the **Manage User Account** window.

Once the **Manage User Account** window has been opened, details can be changed using their associated fields. To manage specific permissions, choose the **Manage Permissions** button and then use the tick boxes to specify which permissions the user should have. Choose **Update** to save the altered permissions and **Update Account** to save the other user detail changes.

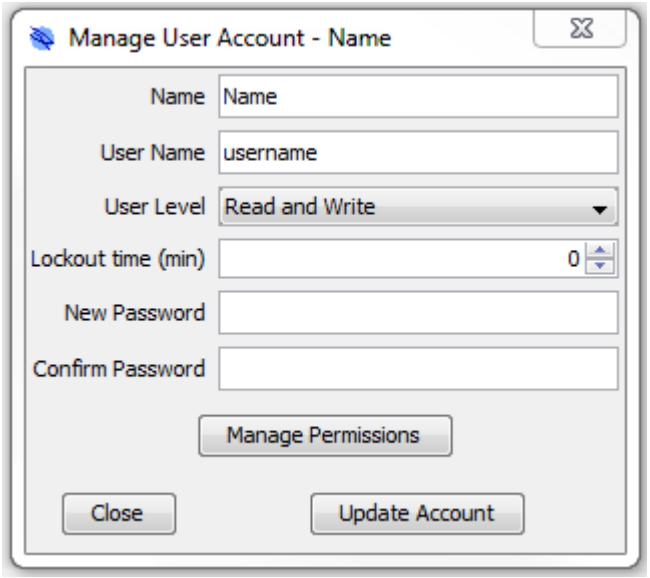


Figure 3: Manage User Account window.

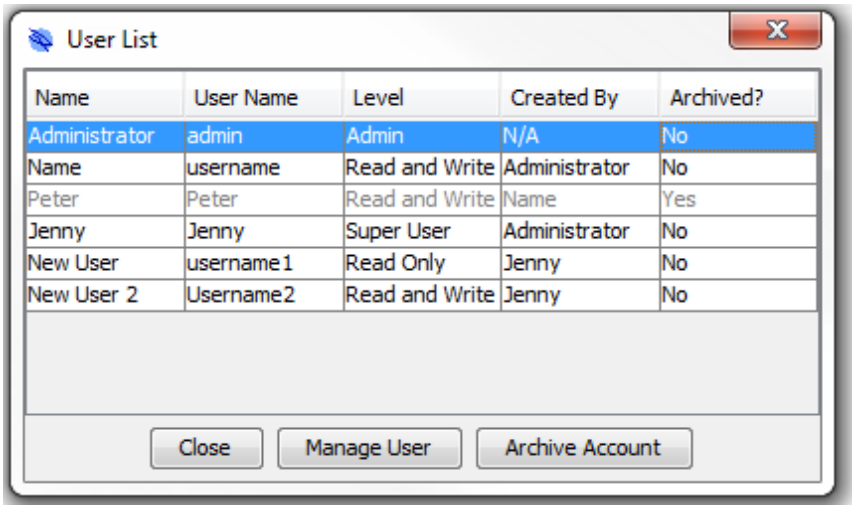


Figure 4: User List window. Grey entries indicate archived users.

Archiving

A user account can be archived by users with permission to manage other accounts. This can be done by selecting **Logged in as [user name] -> User List**, selecting a user from the table and clicking the **Archive Account** button, as seen in Figure 5. Once archived, the account can no longer be used to log in to the software. However, the user name still appears on the submissions made by the user while the account was active and the user information remains, but greyed out, in the **User List**.

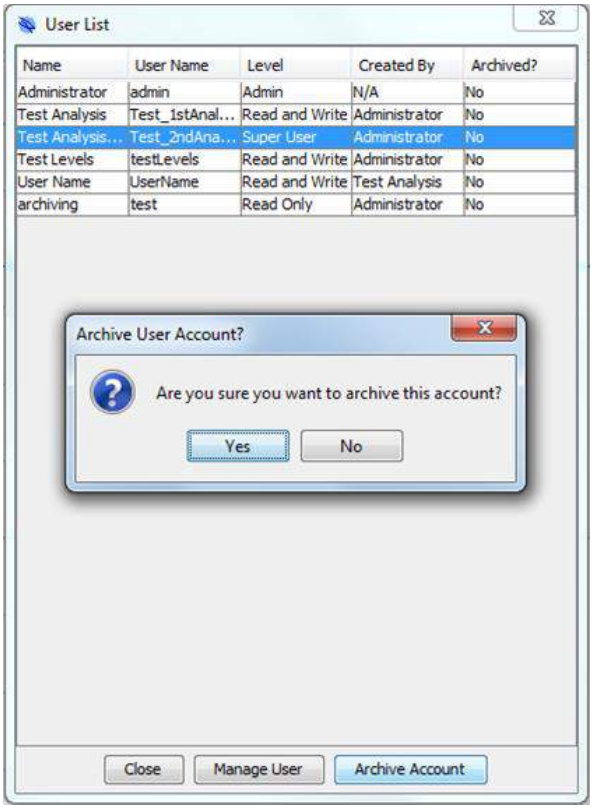


Figure 5: Archiving of a user account.

An archived account can be restored by a managing account by returning to the **User List** window, selecting the user from the list and clicking the **Restore Account** button.

CNV and LOH detection

Workflow strategies

CytoSure Interpret Software has been developed in close collaboration with our customers to provide a flexible solution for labs with a wide variety of analysis requirements. As a result, a number of strategies are available for the use of CytoSure Interpret Software in order to plan and execute interpretation and reporting of CytoSure Array data.

The following sections provide details of 3 such strategies categorised loosely based on the array throughput of the lab. Users are encouraged to assess each strategy in order to select the one most appropriate for their situation.

Pre-requisites

- The following guidance assumes that CytoSure Interpret Software has already been successfully installed. For full details on installation and client configuration, see the Installation Guide (provided with your installation media).
- All strategies assume that a feature extracted file, generated by feature extraction of a TIFF image of a microarray slide, is used as input.
- For full details of the selected strategy, follow the link in the section headings below.

Suggested strategies

- High-throughput — Sample and experiment information is imported into the database in order to automate association between feature extracted files and samples during batch processing. The CytoSure Interpret Software .cgh files generated during batch processing can then be analysed as required.
- Medium-throughput — Batch processing of feature extracted files with manual input of sample information. As with the high-throughput strategy, the CytoSure Interpret Software .cgh files generated for each case during this process can then be analysed as required.
- Low-throughput — Processing individual cases one at a time starting from the feature extracted files.

High-throughput strategy

Overview

The high-throughput strategy (as outlined in Figure 1) involves importing sample and experiment information into the database, including array barcodes, in order to automate association between feature extracted files and samples during batch processing. The CytoSure Interpret Software .cgh files generated during batch processing can then be analysed as required.

- ⊕ Bulk sample data import from automatically generated files prevents data entry errors.
- ⊕ Bulk sample data import reduces hands-on time required for data entry.
- ⊕ Bulk import of expected array barcodes enables automatic sample-feature extracted files to be associated and prevents sample mix-up.
- ⊕ Batch processing eliminates hands-on time required for aberration detection.
- ⊕ Batch processing prevents the need for re-processing and re-entry of sample data.
- ⊖ Additional effort required to setup data import compared to medium- and low-throughput strategies.

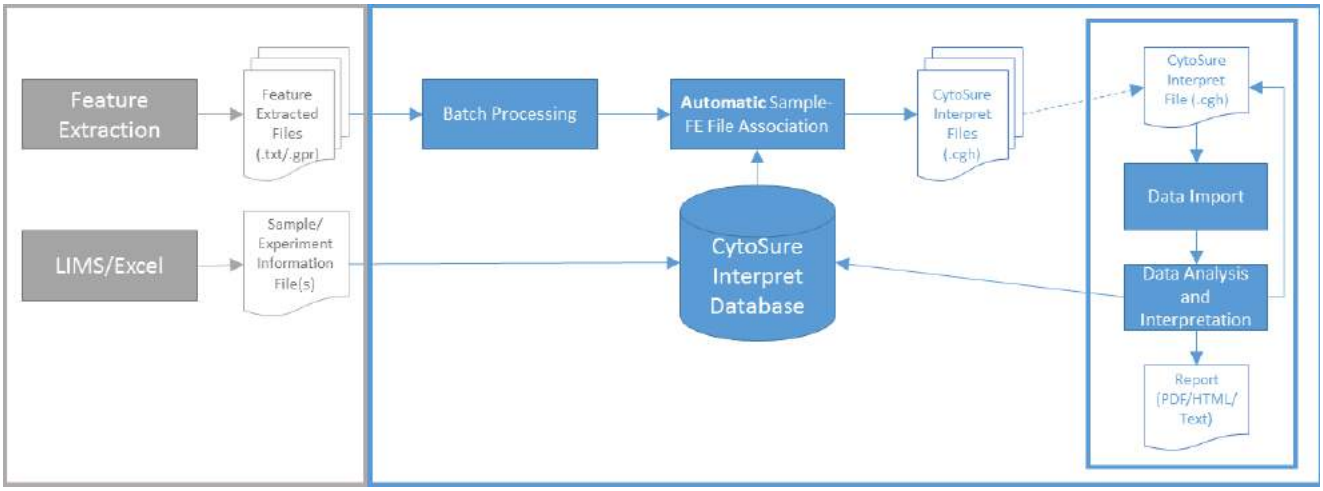


Figure 1: Diagrammatic representation of the CytoSure Interpret Software high-throughput strategy.

Steps (per batch)

A **batch** should be considered as the most convenient way of grouping sample data for analysis. This may equate to multiple slides, a single slide, or a selection of samples from one or multiple slides depending on your circumstances.

1. **Import sample/experiment information** — Importing sample and experiment information, such as sample ID, phenotype and consumable lot numbers, before data analysis, as opposed to after data import (see the Low-throughput or Medium-throughput strategies), reduces hands-on time and the likelihood of data entry errors, whilst also enabling feature extracted files to be linked to sample IDs automatically during Batch processing, further reducing the chance of sample mix-up. See Bulk import sample/experiment information for detailed instructions.

2. **Batch process feature extracted files** — Batch processing involves selecting an analysis protocol, **automatically** assigning sample IDs to the probe data contained within feature extracted files, segmenting the data into regions of equal copy number ready for automated CNV detection and saving all data into CytoSure Interpret Software .cgh files ready for interpretation. For full details of this feature, see the Batch processing section.

3. Then for each case:

a. **Import CytoSure Interpret .cgh file** — The batch processed files are loaded separately into the software to enable manual interpretation and analysis. Methods by which to do this are outlined in Import CytoSure Interpret .cgh files. In particular, option 3 (loading the file via **Database Management)** is the recommended approach in this scenario.

b. **Interpret detected aberrations** — To interpret detected aberrations, it is first necessary to visually observe them. See the View aberrations section for details on the different ways in which to do this. Furthermore, it is possible to assess the aberrations using quality control metrics and annotation tracks (which provide additional external information). See the sections View annotation tracks and Quality metric review for guidance on this. Finally, assessment of this data allows predicted aberrations to be edited, classified and rejected, as discussed in the Edit aberrations and Manual classification sections.

c. **Report results** — Upon completion of the manual checking process, the user has the option to generate a report summarising the sample analysis. For more information, see the Report generation section.

Medium-throughput strategy

Overview

The medium-throughput strategy (as outlined in Figure 2) involves Batch processing of feature extracted files with manual input of sample information. As with the high-throughput strategy, the CytoSure Interpret Software .cgh files generated for each case during this process can then be analysed as required.

⊕ Batch processing eliminates hands-on time required for aberration detection.

⊕ Batch processing prevents the need for re-processing and re-entry of sample data.

⊖ Manual sample data entry requires more hands-on time.

⊖ Manual sample data entry increases likelihood of sample mix-up.

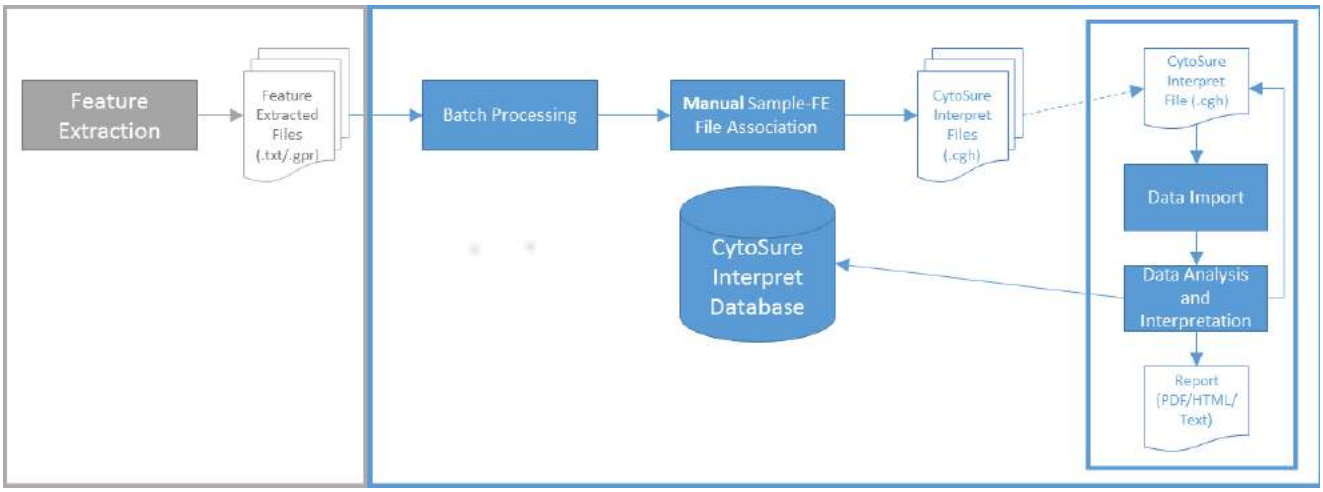


Figure 2: Diagrammatic representation of the CytoSure Interpret Software medium-throughput strategy.

Steps (per batch):

A **batch** should be considered as the most convenient way of grouping sample data for analysis. This may equate to multiple slides, a single slide, or a selection of samples from one or multiple slides depending on your circumstances.

1. **Batch process feature extracted files** — Batch processing involves selecting an analysis protocol, **manually** assigning sample IDs to the probe data contained within feature extracted files, segmenting the data into regions of equal copy number ready for automated CNV detection, and saving all data into CytoSure Interpret Software .cgh files ready for interpretation. For full details of this feature, see the Batch processing section.

2. Then for each case:
    a. **Import CytoSure Interpret Software .cgh file** — The batch processed files are loaded separately into the software to enable manual interpretation and analysis. Methods in which to do this are outlined in Import CytoSure Interpret .cgh files. In particular, option 3 (loading the file via **Database Management)** is the recommended approach in this scenario.
    b. **Interpret detected aberrations** — To interpret detected aberrations, it is first necessary to visually observe them. See the View aberrations section for details on the different ways in which to do this. Furthermore, it is possible to assess the aberrations using quality control metrics and annotation tracks (which provide additional external information). See the sections View annotation tracks and Quality metric review for guidance on this. Finally, assessment of this data allows predicted aberrations to be edited, classified and rejected, as discussed in the Edit aberrations and Manual classification sections.
    c. **Report results** — Upon completion of the manual checking process, it is usually required that a report is generated to summarise the sample analysis. For more information, see the Report generation section.

Low-throughput strategy

Overview

The low-throughput strategy (as outlined in Figure 3 involves processing individual cases one at a time starting from the feature extracted files, rather than pre-processing in bulk via batch processing (see the medium- and high-throughput strategies).

- ⊕ Cases can be analysed immediately as no prior sample information import or pre-processing is required.
- ⊖ Manual sample data entry requires more hands-on time.
- ⊖ Manual sample data entry increases likelihood of sample mix-up.
- ⊖ Aberration detection after data import requires more hands-on time.
- ⊖ Re-processing and re-entry of sample data required if users forget to save after aberration detection.

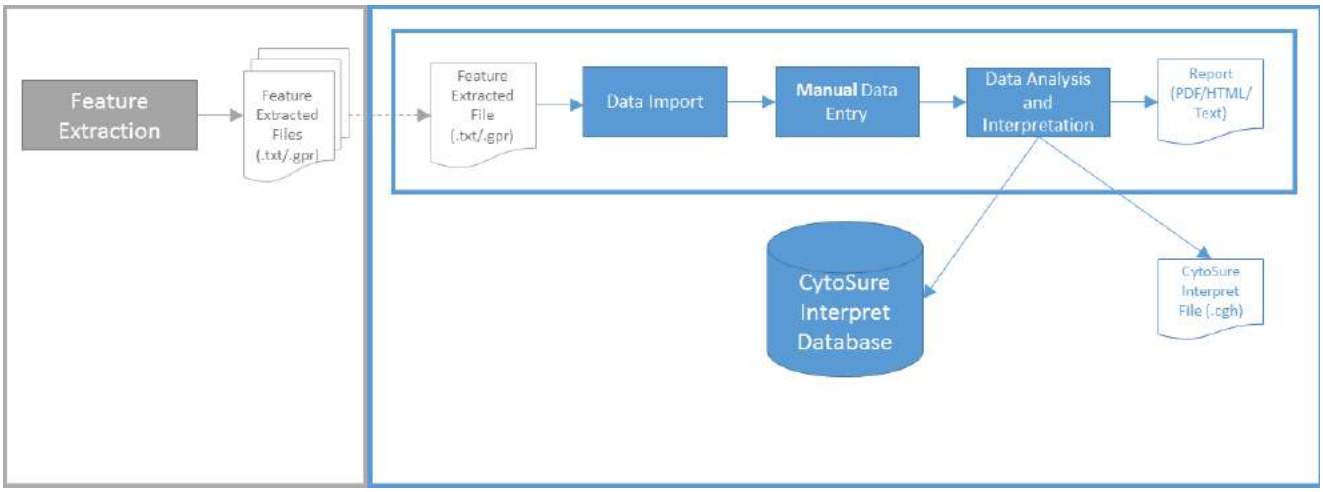


Figure 3: Diagrammatic representation of the CytoSure Interpret Software low-throughput strategy.

Steps (per case):

1. **Load protocol** — Before any analysis is carried out, it is advised that a protocol is loaded specifying a selection of settings for the program. Using a protocol helps to maintain an analysis standard among samples.

2. **Load feature extracted (FE) file** — FE files are individually imported into the software by methods described in the Import feature extracted files section. The simplest method is to use the **Load** icon in the **Accelerate Workflow.**

3. **Sample - FE file association** — Manual entry of sample and experiment information, such as sample ID, phenotype and consumable lot numbers is carried out for each

inputted feature extracted file. This process can be more time consuming and error prone than the automatic approach (see the High-throughput strategy section). See the Individually import sample/experiment information section for instructions on this.

4. **Process data** — Aberration detection is carried out on the individual FE files by segmenting the data into regions of equal copy number. The Individual case processing section describes this process.

5. **Interpret detected aberrations** — To interpret detected aberrations, it is first necessary to visually observe them. See the View aberrations section for details on the different ways in which to do this. Furthermore, it is possible to assess the aberrations using quality control metrics and annotation tracks (which provide additional external information). See the sections View annotation tracks and Quality metric review for guidance on this. Finally, assessment of this data allows predicted aberrations to be edited, classified and rejected, as discussed in the Edit aberrations and Manual classification sections.

6. **Save Data** — Unlike the batch processing methods (High-throughput strategy and Medium-throughput strategy), the aberration detection results must be saved manually in the low-throughput approach. To save to the database select the **Save to Database** button ( ) in the **Sample Tab** on the far right of the **Genomic View** page. To save as a CytoSure Interpret Software **.cgh** file, choose **File -> Save as** or select the save button ( ) in the **Sample Tab**.

7. **Report results** — Upon completion of the manual checking process, a report may be generated to summarise the sample analysis. For more information, see the Report generation section.

Data import

This section describes how to import data into CytoSure Interpret Software in order to enable accurate interpretation of a sample, and is split into two sub-sections:

1. Importing sample and experiment information — detailing the process of adding information relating to the sample and how it has been processed (e.g. Sample ID, Phenotype, Lab Protocol information, lab instrument readings, etc.).
2. Importing array CGH data — detailing how to load array results for analysis and interpretation.

Importing sample and experiment information

Bulk import sample/experiment information

Importing sample and experiment information, such as sample ID, phenotype and consumable lot numbers, before data analysis, as opposed to during data import (see the Low-throughput or Medium-throughput strategies), reduces hands-on time and the likelihood of data entry errors. It also enables feature extracted files to be linked to sample IDs automatically during Batch processing, further reducing the chance of sample mix-up. This technique is utilised in the High-throughput strategy.

Data format

The most efficient method of information import is from a text file, for example, generated from an Excel spreadsheet of samples to be processed or from a LIMS. The file must contain data columns identifiable by any custom separator, for example commas, spaces or tabs (tab-delimited). An example of a spreadsheet version of an information import file is shown in Figure 1.

	A	B	C	D	E	F	G
1	Slide Barcode	Array Position	Array Barcode	Sample ID	Sample Gender	Reference ID	Reference Gender
2	257517510001	1_1	257517510001_1_1	ABC_123	M	Male Reference	M
3	257517510001	1_2	257517510001_1_2	ABC_124	F	Female Reference	F
4	257517510001	1_3	257517510001_1_3	ABC_125	M	Male Reference	M
5	257517510001	1_4	257517510001_1_4	ABC_126	M	Male Reference	M
6	257517510001	2_1	257517510001_2_1	ABC_127	F	Female Reference	F
7	257517510001	2_2	257517510001_2_2	ABC_128	F	Female Reference	F
8	257517510001	2_3	257517510001_2_3	ABC_129	M	Male Reference	M
9	257517510001	2_4	257517510001_2_4	ABC_130	F	Female Reference	F
10	257517510002	1_1	257517510002_1_1	ABC_131	M	Male Reference	M
11	257517510002	1_2	257517510002_1_2	ABC_132	M	Male Reference	M
12	257517510002	1_3	257517510002_1_3	ABC_133	M	Male Reference	M
13	257517510002	1_4	257517510002_1_4	ABC_134	F	Male Reference	M
14	257517510002	2_1	257517510002_2_1	ABC_135	F	Male Reference	M
15	257517510002	2_2	257517510002_2_2	ABC_136	F	Male Reference	M
16	257517510002	2_3	257517510002_2_3	ABC_137	M	Male Reference	M
17	257517510002	2_4	257517510002_2_4	ABC_138	F	Male Reference	M

Figure 1: Example information import spreadsheet for an 8x array. Yellow cells indicate manually inputted data.

The easiest way to create an information import document is to use the spreadsheet templates for 4x or 8x slide formats (which can be provided by OGT on request) as follows:

1. Open the template.
2. Begin by typing the first slide barcode into the yellow cell. The spreadsheet should automatically fill some of the fields and additional yellow cells should appear.
3. Continue to enter data into the yellow cells. Please note that the columns for **Slide Barcode** and **Sample ID** are the only compulsory fields when completing this template.
4. If required, add and complete extra columns for additional information.
5. Once all data has been entered, save a copy of the file in the "Text (Tab-delimited) (*.txt)" format.

6. See File parsers and the import process for information regarding the import of this file into CytoSure Interpret Software.

Alternatively, you can create your own information import text file. At a minimum, the file should include columns for the **Sample ID** and **Array Barcode** (Expected barcode of the array on which the sample has been hybridised), as this enables batch processing to automatically link the feature extracted file to the sample in the database. The exact column titles do not matter, as they are manually linked to data fields when imported into the software. Likewise, the column separator does not matter, as it is specified when the file is parsed into the software.

Custom fields

Custom fields can be created in the **Sample Data** or **Experiment Details** sections of database management to support import and storage of data that is not currently represented in the software.

File parsers and the import process

Once the sample/experiment information file has been generated, it can be imported using a **Parser**. Parsers are user-created configurations for associating data contained in text files with fields in the CytoSure Interpret Software database (e.g. linking column D in Figure 1 with the **Sample ID** field in **Sample Data**, or linking column C with the **Expected Barcode** custom field in **Experiment Details**). If the format of the data file is unchanged, the parser only need be configured once as it is stored and available for all users for subsequent data import.

Creating a parser

To create a parser for the generated sample/experiment information file and to load the data:

1. In the **Database Management** tab, click the the **Attach** () button.
2. If a window appears asking **Would you like to import data using an existing parser?**, click **No**.
3. Select the sample/experiment information import file from your file system and click **Open.**
4. In the **Create File Parser** window (e.g. Figure 2), configure the parser as required for your file format.
    a. Enter a name for the parser in the **Parser Name** field.
    b. Specify the data start line using the **Start from line** box. If you have one header row then this will be **2**.

c. Specify the **Column separator character(s)** of the information import file. If you are using a spreadsheet (e.g. the provided template) that has been saved as a tab-delimited text file, then specify **Tab.**

d. Set the **File Type.** If you are using a template spreadsheet then this will be **One sample per LINE**.

e. Usually users will want to select **Automatically create samples and experiments if required?**; however, if you only want to add information to existing samples and experiments, then leave un-selected.

f. Click the **Update** button in the middle of the screen. The data should be displayed below, without headers. If **One sample per LINE** was selected only the first data row should be displayed, as seen in Figure 2.

g. For each of the cells in the data display:

    i. Click to highlight the cell blue.

    ii. Select the field you would like it to be associated with. For example, if the cell contains a sample gender then select **Basic Field** and choose **Sample Gender** form the drop down. If the field is not present in the **Basic Field** drop down list then select **Custom Field** and choose from this drop-down list. **Expected Barcode**, which is a compulsory field for batch processing, is present in this list. If the required field is not present in either list, then it will be necessary to create a custom field in the **Sample Data** or **Experiment Details** sections of database management.

    iii. Click the **Update** button at the bottom of the screen. Upon doing this, as shown in Figure 2 columns D and E, the cell data should turn green and become surrounded by a border labelled with the field type. If the data turns red then there is a problem, most likely that the data type is not suitable for the field chosen.

h. Once all cells have been associated with a data field, click **Load Data.**

If no errors are detected, the sample/experiment information will have been loaded into the database, although it may be necessary to refresh the **Database Management** tab before the new samples are displayed in the samples table.

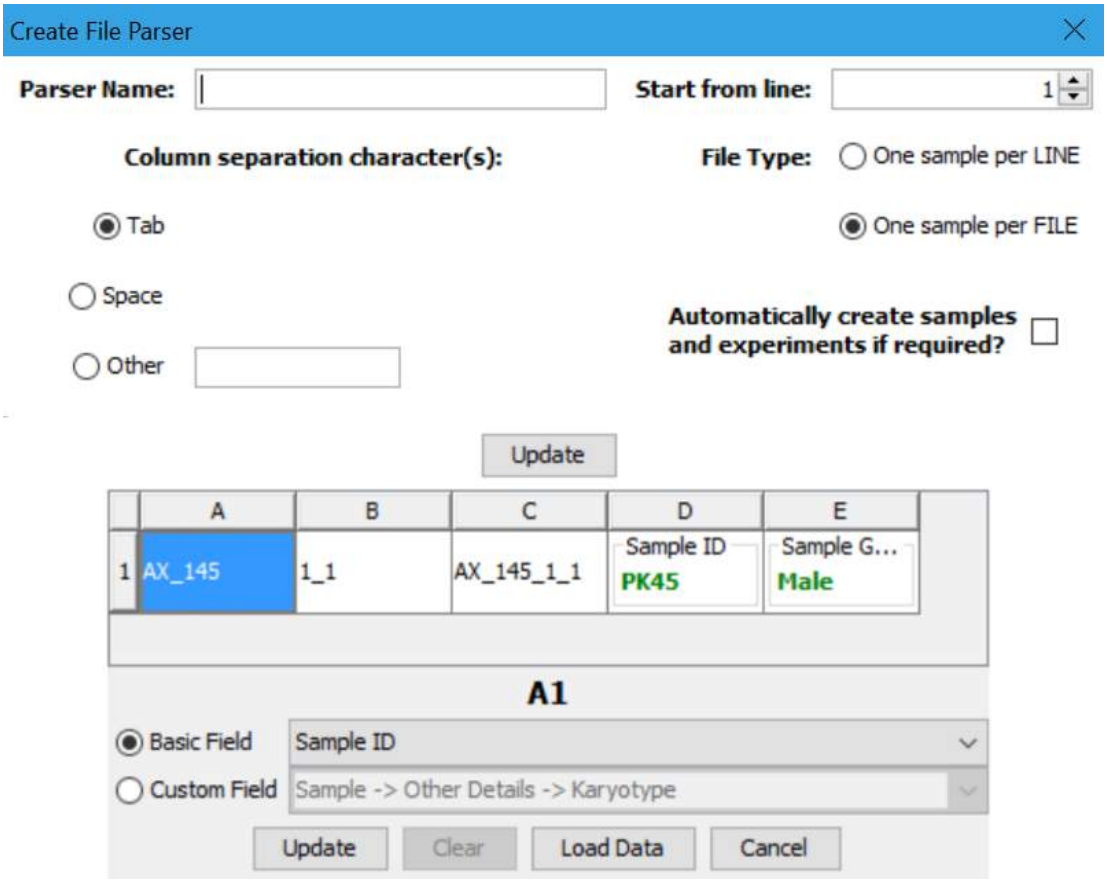


Figure 2: Create File Parser window.

Re-using a parser

Sample/experiment files of the same format can subsequently be imported as follows:

1. In the **Database Management** tab, click the **Attach** () button.
2. When a window appears asking **Would you like to import data using an existing parser?**, click **Yes**.
3. Select the previously created parser from the list in the **Select Parser** window.
4. When asked **Would you like to preview the file before importing data?**, click **No**. (Click **Yes** if you would like to see the contents of the file to be selected and have the option to make changes to the parser, then follow Creating A Parser from step 3.)
5. Select the sample/experiment information import file from your file system and click **Open**.

If no errors are detected, the sample/experiment information will have been loaded into the database, although it may be necessary to refresh the **Database Management** tab before the new samples are displayed in the samples table.

Individually import sample/experiment information

Importing sample and experiment information, such as sample ID, phenotype and consumable lot numbers, is carried out manually in the Low-throughput and Medium-throughput strategies. It can be done before data analysis has begun using the **Database Management** tab or during data import using the **Sample Details** window.

Before data analysis

To add samples and experiment information before data analysis has begun, navigate to the **Data Management** tab. Choose **Sample Data** from the central panel (Figure 3) followed by the **Add/Create** icon ( ). This will create a new sample with an automatically assigned sample ID, details of which will appear in the **Sample Information** section on the right side of the page (Figure 4). Use the entry boxes in this section to add any required sample details, and if necessary change the sample ID. To save, click the **Save** icon ( ) to the left side of the sample ID in the sample tab at the top right of the **Database Management** tab.

To add an experiment and its details, choose **Experiment Details** from the central panel (Figure 3) followed by the **Add/Create** icon ( ). This will bring up a streamlined **Sample Details** window, as seen in Figure 5, from which to choose the Sample ID and Reference ID for the experiment (the samples required must be in the database first!). Once chosen, the experiment will be created, loading the **Experiment Information** section on the right side of the page. Use the entry boxes in this section to add any required experiment details. To save, click the **Save** icon ( ) to the left side of the sample ID in the sample experiment tab at the top right of the **Database Management** tab.

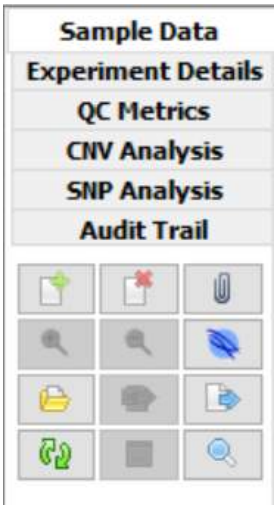


Figure 3: Central panel of the Database Management tab.

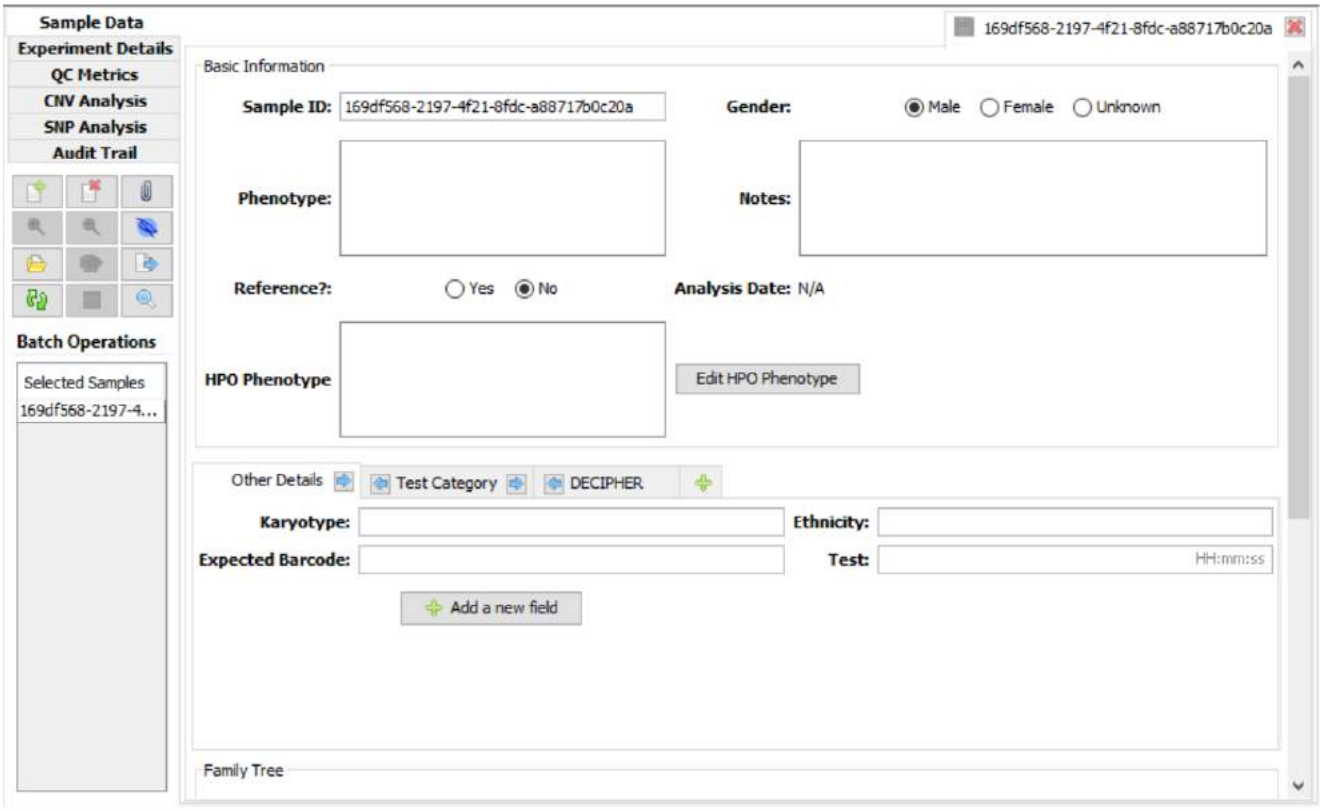


Figure 4: Central panel and Sample Information section of the Database Management tab.

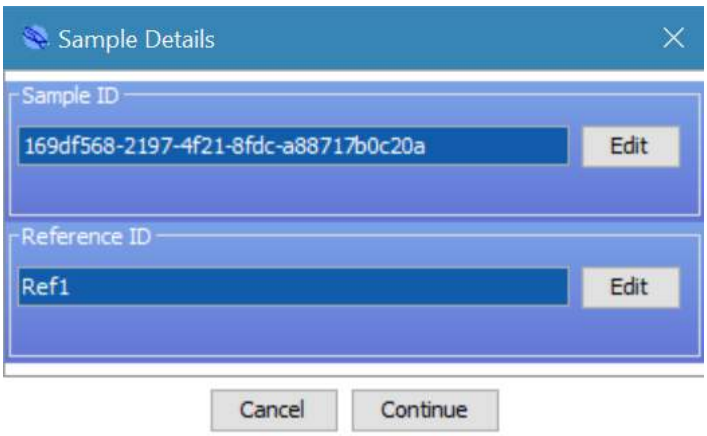


Figure 5: Sample Details window for the selection of Sample and Reference IDs during creation of the experiment.

During data import

Feature extracted files can be imported individually or through a batch process (Individual case processing or Batch processing).

Batch processing

For the medium-throughput strategy (Medium-throughput strategy), batch processing is used. To use this strategy, follow the steps on the Batch processing page, including the instructions to manually specify files. Once files are present in the **Files to be processed** table, double-click on each row in the list to bring up the **Sample Details** window, as seen in Figure 6. This window is used to specify the sample and reference ID as well as add sample data and experiment details. As you type into the **Sample ID field**, any matching Sample IDs within the database will be displayed for selection. Alternatively, type in the ID required for a new sample and click **Create New Sample**. Upon doing this, a drop-down bar labelled **Sample Data** will appear**,** from which some sample details can be specified (Figure 7). Carry out the same process for the reference ID and then the experiment details in the **Experiment Details** section as required. Select **Continue** to save the data.

> It is important to remember how the sample/reference DNA is labelled (Cy3/Cy5). This information will be displayed within the **Experiment Details** panel at the bottom of the screen.

Alternatively, if no data regarding the samples has been added to the database, for convenience during validation (but not common use) choose the **Process** button in the **Batch Processing** window. This will bring up the **Missing Details** window. The **automatically generate new samples** option would be used if you wish the software to randomly generate sample IDs. Alternatively, the option **Use barcode as sample ID** can be used to create new samples using the barcode as a sample ID.

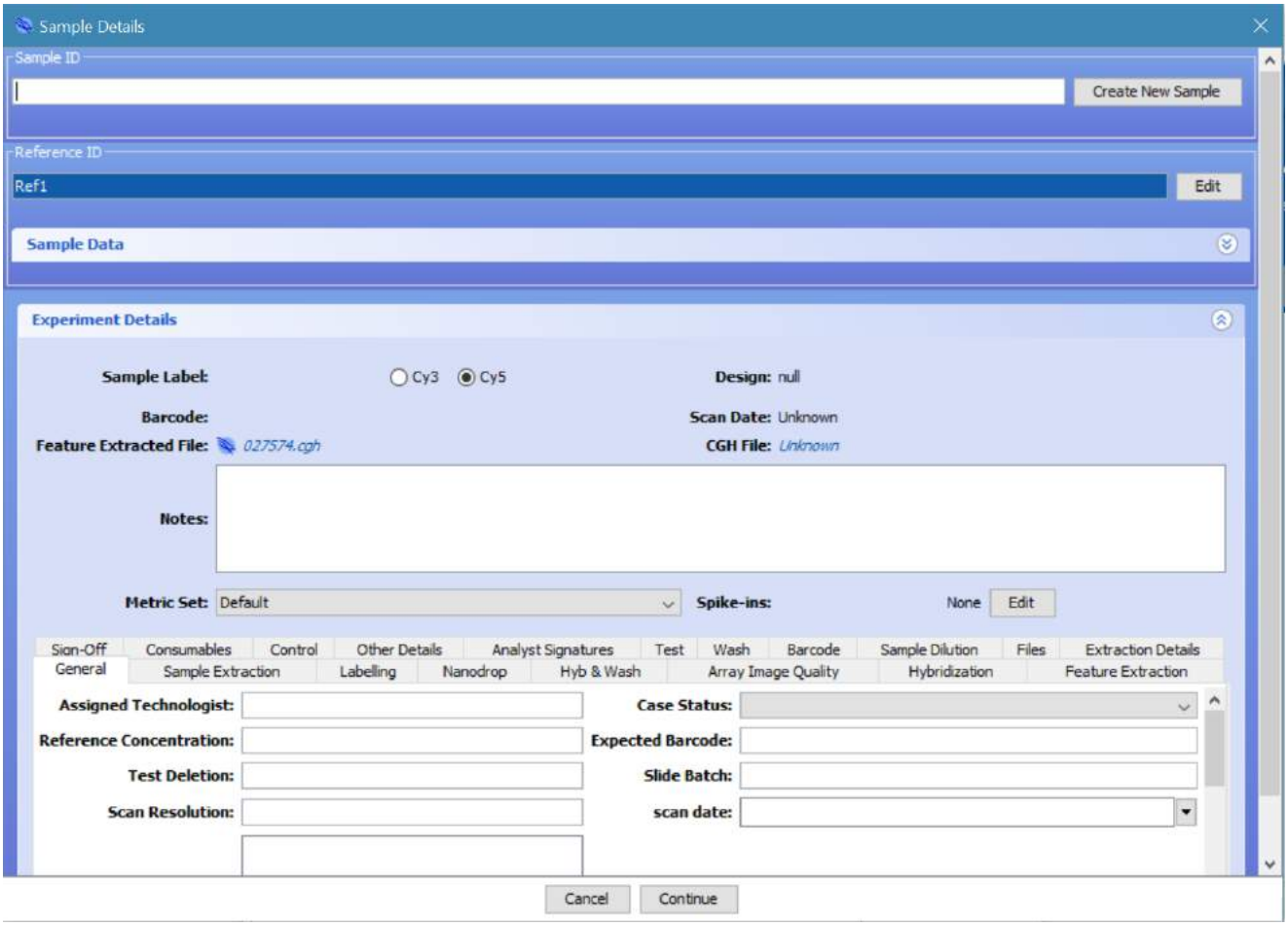


Figure 6: Full Sample Details window

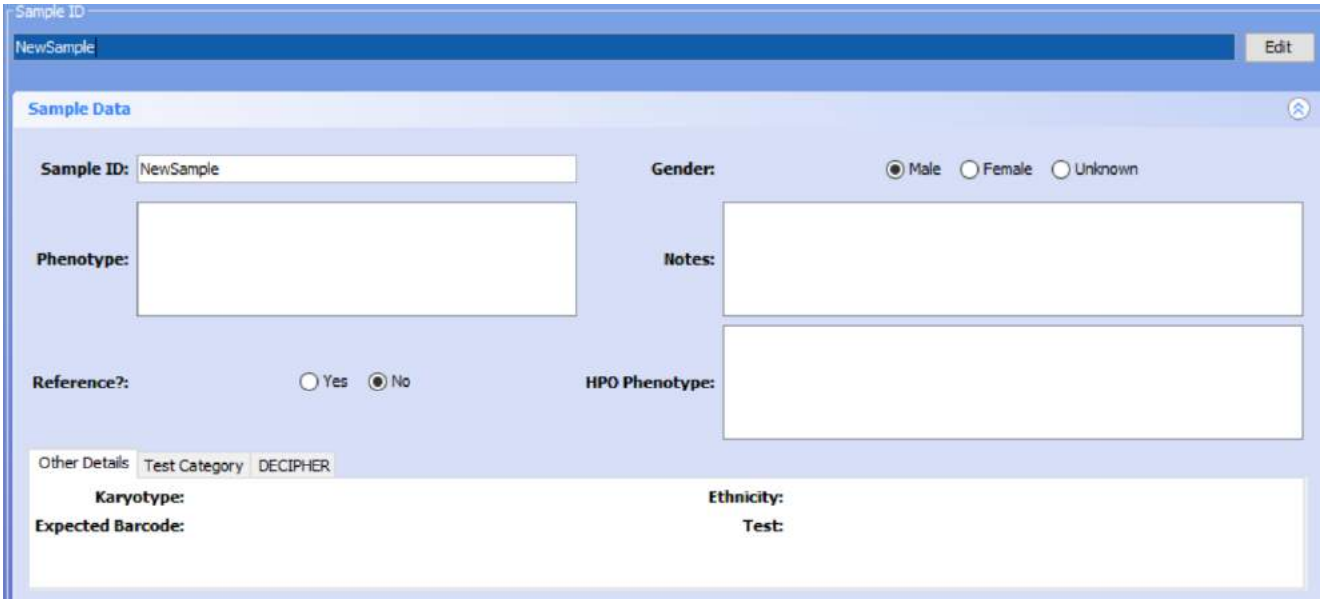


Figure 7: Sample Data input options in the Sample Details window.

Individual case processing

For the low-throughput strategy (Low-throughput strategy), individual case processing is used. Using this strategy, feature extracted files can be loaded by using the **Load** button in the **Workflow Bar** (Figure 8) and subsequently selecting **Feature Extracted Files** or by choosing **File -> Import** followed by the file type of interest. Either of these methods will load the **Sample Details** window (Figure 6). This window is used to specify the sample and reference ID as well as add sample data and experiment details. If the sample has already been created in the database, then type the sample ID into the relevant box. Alternatively, type in the box the ID required for a new sample and choose **Create New Sample.** Upon doing this, a drop-down bar will appear labelled Sample Data, from which some sample details can be specified (Figure 7). Carry out the same process for the reference ID and then the experiment details in the **Experiment Details** section as required. Select **Continue** to save the data.

Figure 8: Load icon in Workflow Bar.

Importing array CGH data

Import CytoSure Interpret Software .cgh files

CytoSure Interpret Software .cgh files, which are generated during processing of feature extracted files, can be opened for viewing and interpretation in a number of different ways:

1. In the menu bar choose **File -> Open** (as seen in Figure 9). Then choose a **.cgh** file from the File Explorer.

2. Locate the **Workflow Bar,** as seen in Figure 10. Click on the **Load** arrow, and a **Data format selection** window will appear, as seen in Figure 11. Choose the **Batch Processed Files** button from the window and choose the appropriate file from the File Explorer.

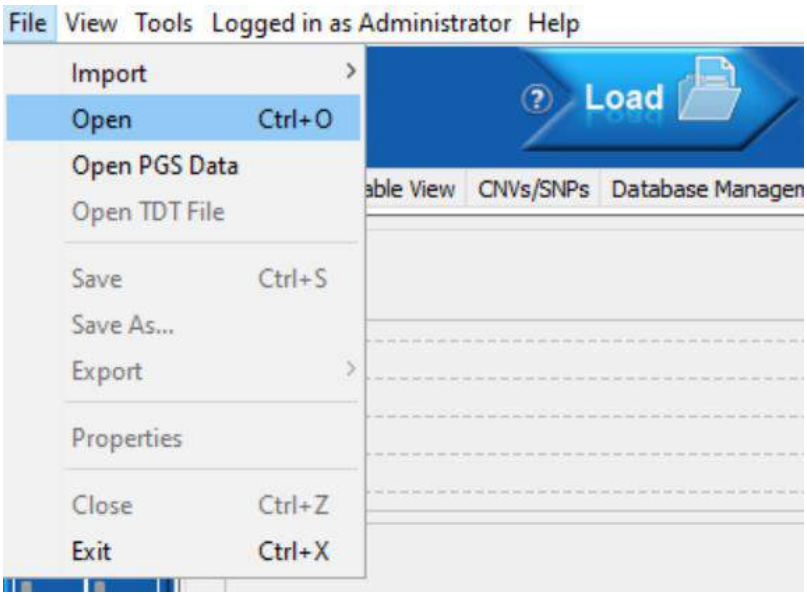


Figure 9: Opening a .cgh file using the File -> Open option.

Figure 10: Workflow Bar.

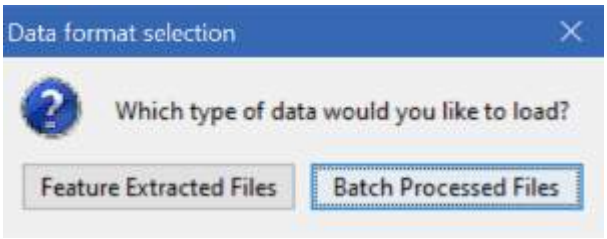


Figure 11: Data format selection window.

3. Alternatively, open a **.cgh** file using the **Database Management** tab. First, open the tab and select the relevant sample from the list. In the button grid**,** locate and click the **CytoSure Interpret Software** icon (which displays **Open CytoSure Interpret File** when hovered over). These actions can be seen in Figure 12.

> Note that this will only result in a .cgh file being opened if the path to the file has been set for the sample's default experiment, and that file exists.

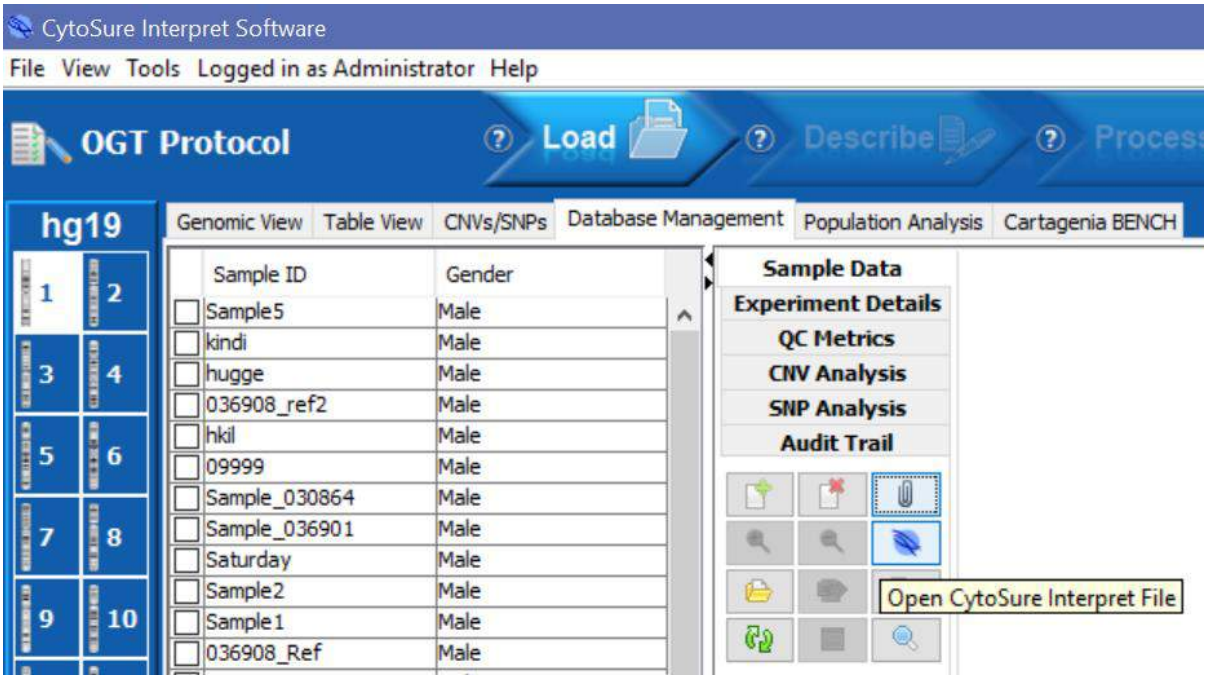


Figure 12: Opening a .cgh file using the Database Management tab.

4. If batch processing has been carried out, as described in Batch processing, then it is also possible to open the **.cgh** files from the Batch Processing window (Figure 13). When completed, the files in the **Files to be processed** window will display **Completed** in their **Status** column. To open the associated **.cgh** file, either double-click on the row in the table or single click on the row (to highlight it blue) and then press the **Load into workflow** button.

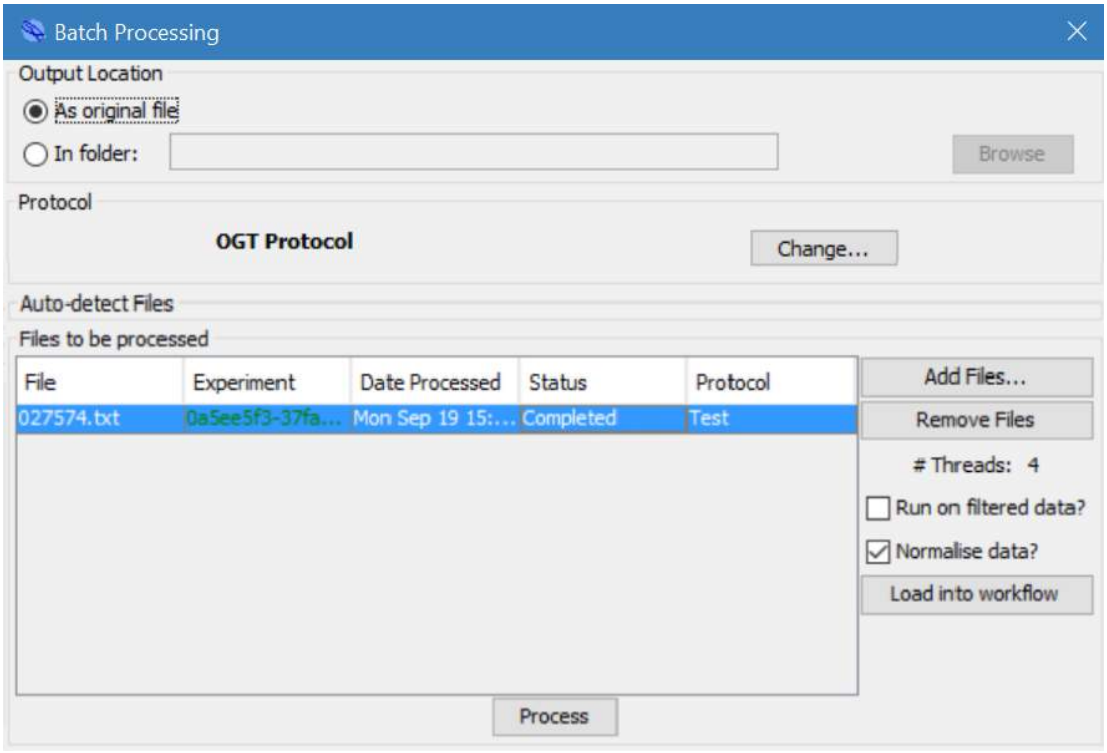


Figure 13: Batch Processing window.

5. It is also possible to drag and drop .cgh files from Windows Explorer into CytoSure Interpret Software in order to load them for analysis. Please note that, if this method is used, the data will be loaded outside of Workflow mode.

Import feature extracted files

Using workflow mode

The workflow mode is accessed via the **Accelerate Workflow Bar** at the top of the window (Figure 14). To import a feature extracted file:

1. Click the **Load** icon on the **Workflow bar**.

2. In the dialog window (Figure 15) click the **Feature Extracted Files** button to load a feature extracted .txt or .gpr file. Alternatively, click the **Batch Processed File** button to load data previously analysed by CytoSure Interpret Software (.cgh files).

> For more information on batch processing and loading batch processed files, see the Batch processing section.

3. After clicking on the appropriate button, a browser window will appear which can be used to navigate to the appropriate location in your file system. After highlighting the required data file, click **Open.**

4. At this point, if the design hasn't been opened previously within CytoSure Interpret Software, the software will ask for a design file (.xml). In this instance, you need to browse to the location for the design file, highlight the specific file and click on **Open**. This only needs to be performed once upon analysing a new design for the first time.

   All design files for CytoSure microarrays can be found on the CD supplied with the arrays.

5. The **Sample Details** window should now open, from which sample information and experiment details can be added for the input file. A new Sample ID can be created by typing within the **Sample ID** field. As you type, any matching Sample IDs within the database will be displayed. Click on the displayed ID to use this Sample ID or select **Create New Sample**. A Reference ID must also be chosen or created. For more information on using the **Sample Details** window, please refer to the Individually import sample/experiment information section.

6. It is important to remember which way round the sample/reference DNA is labelled (Cy3/Cy5). This information will be displayed within the **Experiment Details** panel at the bottom of the screen. Click the **Continue** button to save the sample information and continue.

7. The file is loaded automatically into CytoSure Interpret Software.

8. Refer to the Low-throughput strategy page for guidance on how to continue analysis after import.

Figure 14: Accelerate Workflow bar.

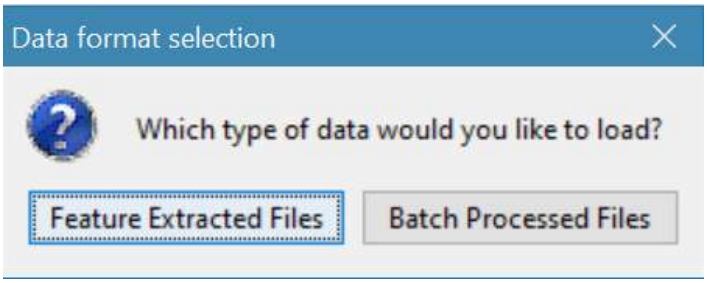


Figure 15: Data format selection dialog window.

Using other methods

Outlined below are alternative methods to load feature extracted files which do not utilise the **Accelerate Workflow bar**.

1. In the toolbar, choose **File -> Import** followed by **Agilent Format** or **GenePix Format** to load feature extracted .txt or .gpr files respectively (Figure 16). A browser window will appear which can be used to navigate to the appropriate location in your file system. After highlighting the required data file, click **Open.**
2. Drag and drop one or more feature extracted files from Windows Explorer into CytoSure Interpret Software.

3. Feature extracted files can also be loaded during batch processing. For more details, see the Batch processing section.

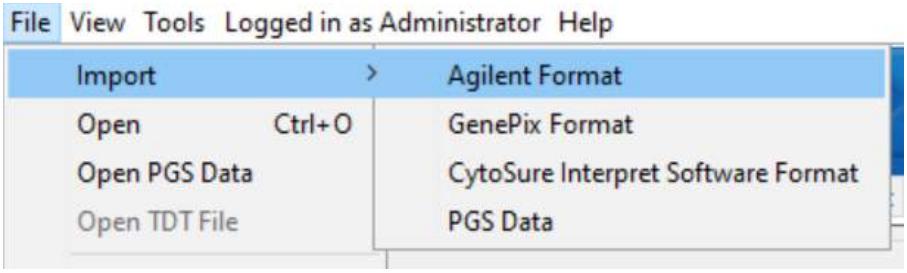


Figure 16: Importing Agilent Format feature extracted files using the CytoSure Interpret Software toolbar.

Data processing

Batch processing

Batch processing (which is utilised in the High-throughput and Medium-throughput strategies) is a method in which large quantities of feature extracted files can be analysed with little manual intervention. To set up batch processing, a group or a folder of feature extracted files is specified, along with the protocol and output location. The processing uses a circular binary segmentation approach (Olshen *et al*., 2004*) to identify regions of equal

copy number for aberration detection. All output data will be saved into CytoSure Interpret Software .cgh files ready for interpretation. This method eliminates the hands-on time required for aberration detection and prevents the need for re-processing and re-entry of sample data.

If following the high-throughput approach, sample and experiment data (for the samples to be batch processed) is bulk imported into the database via a parser utility (Bulk import sample/experiment information) before batch processing is set up. Therefore, when batch processing begins, the system can search the CytoSure Interpret Software database for the associated sample and experiment information, automatically associating feature extracted data with its associated information, before processing the files (and any files which are subsequently added to the folder) without manual intervention. If following the medium-throughput approach, the Sample IDs are **manually** assigned to the probe data contained within feature extracted files.

To carry out batch processing on feature extracted files:

1. If following the high-throughput approach, ensure all sample and experiment information has been imported into the database. At a minimum, this data should contain **Array Barcode** and **Sample ID** for every sample. See Bulk import sample/experiment information for further details. If following the medium-throughput approach, all sample and experiment information should also have been added to the database, but using the individual import options not the bulk import ones. For more details on this see the Individually import sample/experiment information section.

> You can see what data is present in the database by looking at the **Database Management** tab of the main CytoSure Interpret Software screen.

2. Select **Tools -> Batch Processing** to open the **Batch Processing** window, as seen in Figure 1.

3. Select the output location, into which the software will save all CytoSure Interpret Software .cgh files. If **As original file** is selected, the .cgh file will be saved in the same location as the input file is located.

4. If necessary, change the protocol required by clicking **Change** in the protocols box. This will open the **Protocols** tab of the **Options** menu.

The **Run on filtered data?** and **Normalise data?** tick boxes towards the lower right of the **Batch Processing** window (Figure 1) cannot be edited in this window. The **Run on filtered data?** box is checked if the **Apply probe filters during Aberration Detection?** option in the **Probe Filtering** tab in the **Options** window has been selected. If this option is not selected, the filters will just remove probes from the view rather than from the results. The **Normalise data?** box is checked if the **Automatically normalise during protocol?** option in the **Normalisation** tab in the **Options** window has been selected.

5. Specify which files to process and begin processing them. Do this by either:
    a. Using the **Auto-detect Files** section to automatically find all feature extracted files in a selected folder. See Auto-detect files for more information.
    b. Adding specific files to the **Files to be processed** box and process them following the guidelines in Manually specify files.

6. Processing the files listed in the **Files to be processed** window will continue as long as CytoSure Interpret Software is open, even when the **Batch Processing** window is closed. To see the status of file processing, open the **Batch Processing** window and locate the **Status** and **Date Processed** column in the **Files to be processed** table (which will display **Completed** and the completion date/time respectively when the analysis is complete).

7. Once a feature extracted file analysis has been completed, it is possible to load the associated CytoSure Interpret Software .cgh file into the program through the **Batch Processing** window. This can be done as follows:
    a. Highlight the row(s) of the file(s) required by clicking on it/them in the **Files to be processed** table. Highlight multiple rows at once by holding down the **shift** key whilst clicking on the rows.
    b. Double-click on the highlighted section or press the **Load in workflow** button at the bottom right of the window. Please note that the files can take a few seconds to load.
    c. If the **Load into workflow** button is used, the **Sample Details** window will open, which allows you to enter any additional sample, reference or experiment details. Click **Cancel** to stop the program opening the data, or **Continue** to open the data (even if additional data has not been entered).

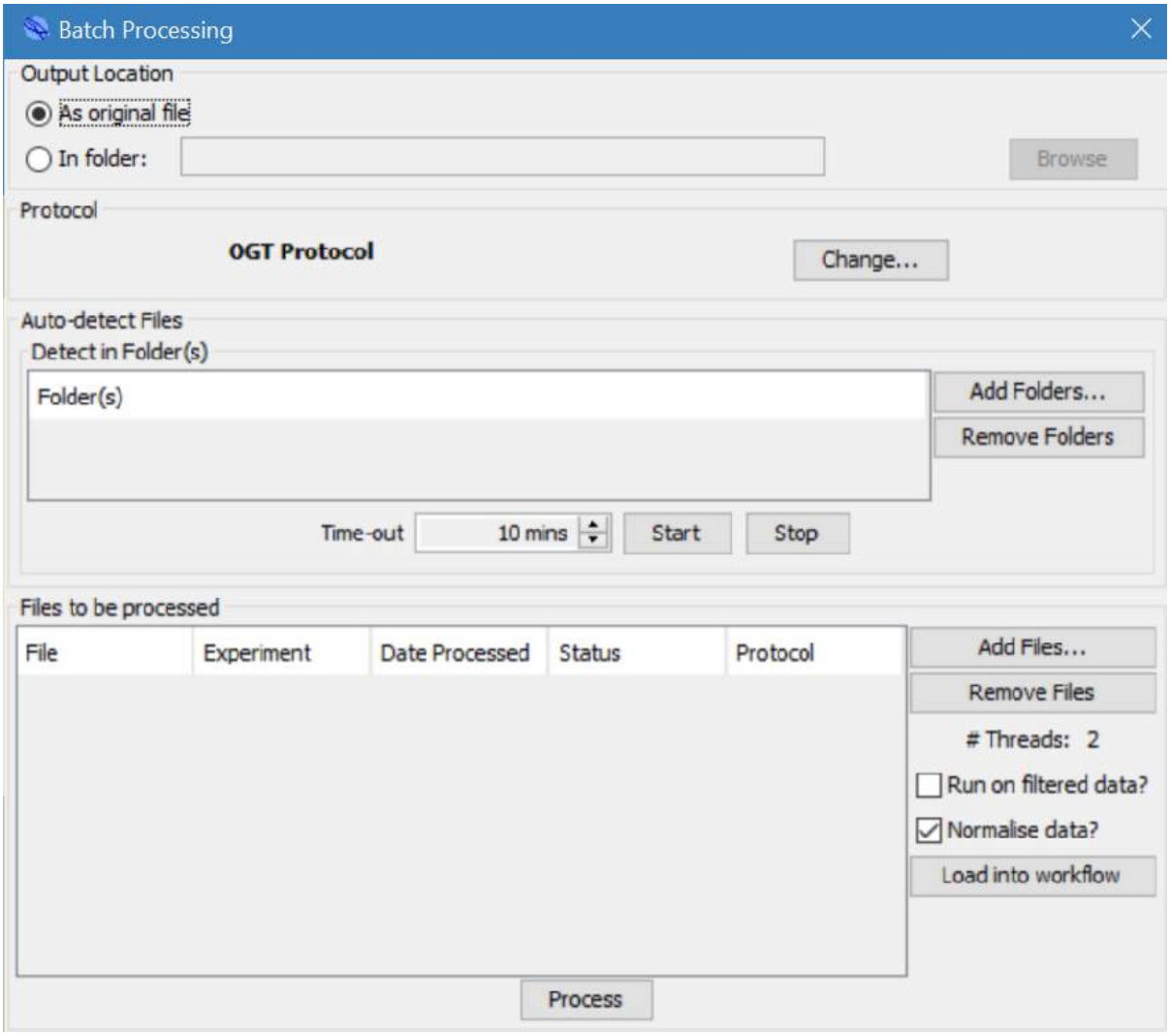


Figure 1: Batch Processing window.

Auto-detect files

If you want the system to automatically detect files from a folder, select the folder/s in the **Auto-detect Files** section of the **Batch Processing** window (See Figure 1).

1. First, select the folder(s) containing the required feature extracted files using the **Add Folder(s)** button. They should now be displayed in the **Folder(s)** table.

2. Alter the **Time-out** number as required. The **Time-out** option determines how long the system will look for files after **Start** is pressed. Any **.txt** files added to the selected folder/s in this time should be identified and added to the **Files to be Processed** table. The search for files will continue even when the **Batch Processing** window is closed, as long as **Time-out** has not been reached. Processing of the files, on the other hand, will continue when the **Batch Processing** window is closed and past the end of **Time-out**. Once started, the progression towards **Time-out** can be seen as a number counting

upwards (towards the set **Time-out** figure) in the location of the **Start** button (in the **Auto-detect Files** section).

3. Press **Start** to initiate the search for all **.txt** files in the selected folder/s. Each file found is added to the **Files to be processed** table and automatically processed.

> When choosing a folder in the **Auto-detect Files** section, please ensure that all **.txt** files in the chosen folder are valid feature extracted files. The software will attempt to process all text files and display the message **Unable to parse file.txt** if a file is invalid (e.g. If it isn't a feature extracted file).

4. Upon pressing **Start,** the system should display a message (see Figure 2), which asks how you would like to associate the generated experiments to a sample. For details on these options, see the Associate feature extracted files with samples section.

5. Press **Stop** at any time to stop the system auto-detecting files.

6. The files which have been found can be seen in the **Batch Processing** window (even if it has been closed and re-opened), in the **Files to be processed** table. This table will also display the status of processing and, if appropriate, the date and time in which processing was carried out.

7. Closing the program will stop **Batch Processing**, clear all files to be processed and deselect any folders.

Manually specify files

1. Press **Add Files** to choose feature extracted files from the File Explorer, or drag files from the File Explorer and drop them in the **Files to be processed** table.

2. The files chosen should be listed in the **Files to be processed** table. Check that the files contain the correct details.
    a. If any fields are red, this indicates that additional data is required. If you double-click on the field, the **Sample Details** window will open, from which sample and experiment details can be manually added. If it is necessary to do this, it is likely

that the bulk import of information into the database was not completed successfully.

b. If you want to use a different protocol for different files, then click on the **Protocols** field in the **Files to be processed** table to bring up a drop-down menu of available protocols. Choose the required protocol.

c. If required, remove files from the table by pressing **Remove Files.**

3. If following the medium-throughput approach, associate the FE files with Sample IDs by double-clicking on the associated row in the **Files to be processed** table. Further details are given in the Associate feature extracted files with samples section.

4. Once all required files are present in the **Files to be processed** table, click the **Process** button at the bottom of the window.

5. If following the high-throughput approach and Sample IDs have not yet been assigned, then the system should display a message (see Figure 2), which asks how you would like to associate the generated experiments to a sample. For details on these options, see the Associate feature extracted files with samples section.

Associate feature extracted files with samples

High-throughput strategy

When samples have been specified using the Auto-detect files section, upon pressing **Start,** the system should display a message (Figure 2) which asks **How would you like to associate the generated experiments to a sample?** or **Experiment details are missing from some cases, how would you like to proceed?**. This option box is also displayed when the **Process** button is chosen on the Manually specify files section.

As sample and experiment information for the added files should already be present in the database, select **Match barcodes to existing sample information**. Then pick from the list which data type/field (from the database information) should be matched to the slide barcode of the files to be processed (Note: This should match the data type/field selected for the slide barcode column when creating the file parser for the bulk import, as described in Bulk import sample/experiment information). The **Status** column of the **Files to be processed** table will display **Error** if the software was not able to detect the sample associated with the barcode of a feature extracted file.

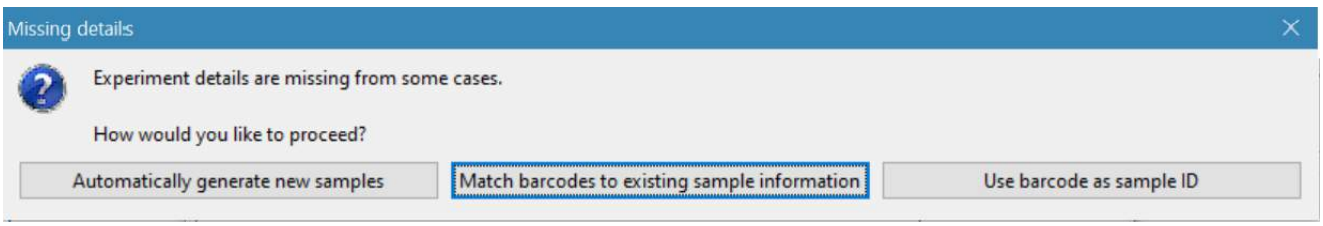


Figure 2: Missing details message window

Medium-throughput strategy

The standard method for the medium-throughput strategy is to add files for processing using the Manually specify files section. Once files are present in the **Files to be processed** table, double-click on each row in the list to bring up the **Sample Details** window, which is used to specify the Sample ID and Reference ID. Alternatively, if no data regarding the samples has yet been added to the database, for convenience during validation (but not common use) choose the **Process** button to bring up the **Missing Details** window (Figure 2). The **Automatically generate new samples** option would be used if you wish the software to randomly generate sample IDs. Alternatively, the option **Use barcode as sample ID** can be used to create new samples using the barcode as a sample ID.

*CBS detection: Olshen, A.B., Venkatraman, E.S., Lucito, R. and Wigler, M. (2004) 'Circular binary segmentation for the analysis of array-based DNA copy number data', *Biostatistics, vol. 5(4), October, pp. 557-72.*

Individual case processing

Following the import of individual feature extracted files and their associated information into CytoSure Interpret Software (Import feature extracted files; Individually import sample/experiment information), individual case processing can be carried out. During the loading of feature extracted files, the probes are mapped to the genome. Once loaded, the data is normalised according to the protocol settings. Processing refers to the following step in which aberrations are identified using Circular Binary Segmentation (Olshen *et al*., 2004*).

Processing can be initiated in two ways:

1. Click **Process** in the **Accelerate Workflow Bar** (Figure 3), provided the data has been loaded using the workflow bar.

Figure 3: Accelerate Workflow Bar.

2. Choose **Tools -> Identify Aberrations** from the toolbar. A window will be displayed asking if you want to carry out Circular Binary Segmentation on your sample, choose **Yes** (Figure 4)**.**

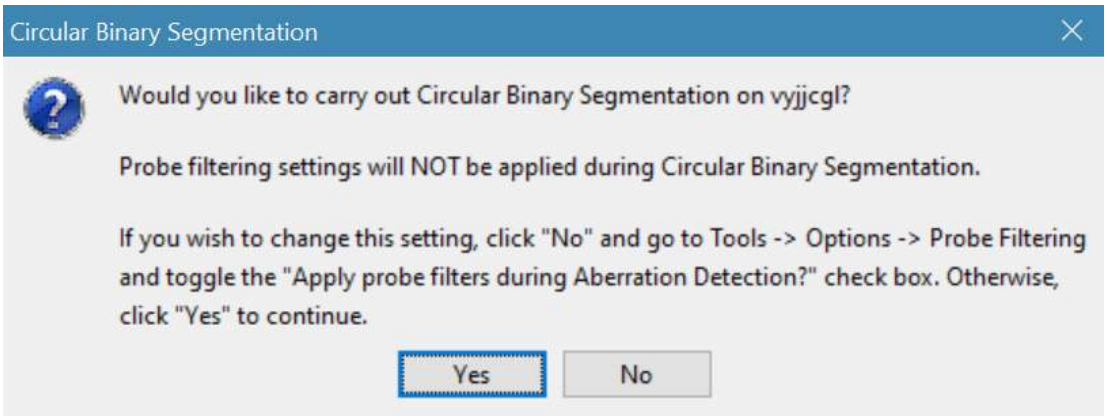


Figure 4: Circular Binary Segmentation window.

While the data is being analysed, the **CBS Progress** window is displayed. Depending on the speed of the computer and the size of the data set being analysed, this can take a couple of minutes to complete. The analysis time can be reduced by selecting **Add a new thread** (Figure 5).

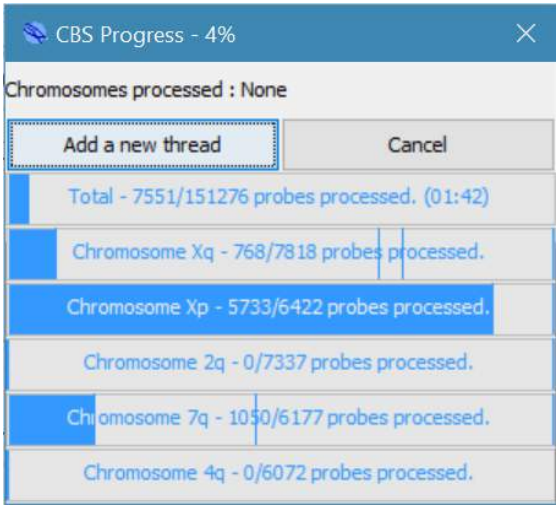


Figure 5: CBS Progress window.

After processing, the **CBS Progress** window disappears and the display re-loads.
The **Aberrations** panel, at the bottom right of the page, will be populated.

*CBS detection: Olshen, A.B., Venkatraman, E.S., Lucito, R. and Wigler, M. (2004) 'Circular binary segmentation for the analysis of array-based DNA copy number data', *Biostatistics, vol. 5(4), October, pp. 557-72.*

Trio analysis

If parental data is available, it can be loaded into the software and linked to the sample, providing inheritance information for detected aberrations. To link the sample to parental data, click **File** -> **Properties**, select the relevant tab and choose the appropriate datasets in the **Mother** and **Father** drop-down boxes, as seen in Figure 6.

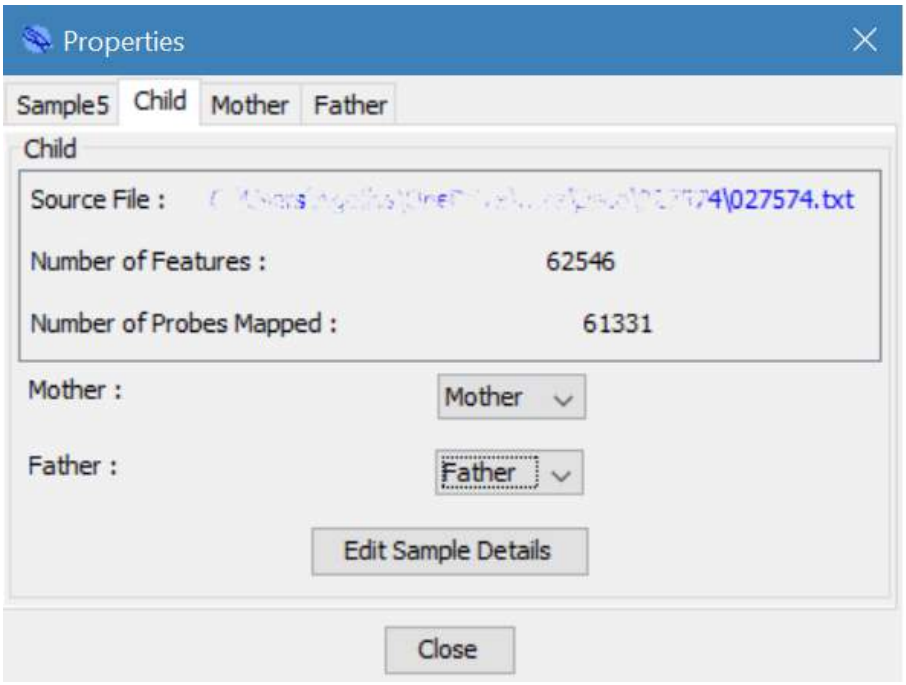


Figure 6: Properties dialog box containing Mother and Father sample drop-down boxes.

If aberrations have been detected in the datasets, the software will attempt to attribute gains and losses in the sample to the mother and/or father samples on the following basis:

- If there are no maternal or paternal aberrations overlapping the aberration region, **Inheritance** will be set to “De Novo” and **Inheritance %** will be set to 100%.
- If only one parent has an aberration overlapping the aberration, **Inheritance** will be set to either “Maternal” or “Paternal” as appropriate, and **Inheritance %** will be set to the percentage of the aberration region which is covered by the overlapping parent aberrations.

If both parents have aberrations overlapping the aberration region, **Inheritance** will be set to “Biparental” and **Inheritance %** will be set to the percentage of the aberration region which is covered by either a maternal or paternal aberration.

Feature extraction

Feature Extraction, the process of calculating probe intensities from microarray image files, is usually carried out by software provided with the microarray. However, CytoSure Interpret Software can perform feature extraction on TIFF (Tag Image Format) images generated from a number of different scanners via the Feature Extraction window (**Tools -> Feature Extraction**) (Figure 7). For full information on how to take advantage of this functionality, please refer to the Feature Extraction User Guide (**Help -> Feature Extraction Guide**).

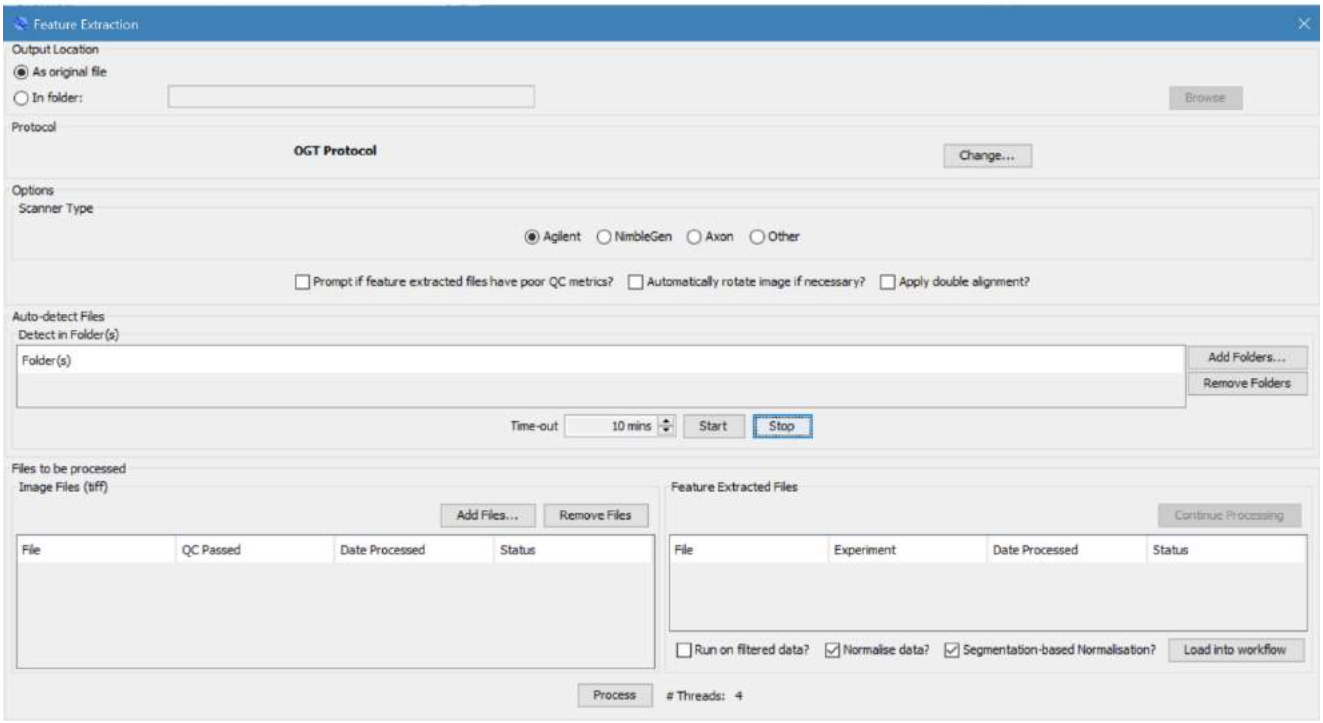


Figure 7: Feature Extraction window.

Data interpretation

Following the detection of aberrations, both the CNVs and any LOH regions may be visually reviewed and then classified. The process of review may include visual inspection (View aberrations), quality control metric checks (Quality metric review) and review of associated annotations (View annotation tracks). Following a thorough review, aberrations can then be classified (initially and/or finally) following the guidelines in the Manual classification section. Refer to the following section for guidance on each of these steps.

View aberrations

Once a CytoSure Interpret Software **.cgh** file has been opened, there are three displays which can be used to view the data: **Genomic View**, **Table View** and **CNVs/SNPs**. These views can be opened by switching between the main tabs, as seen in Figure 1 (near centre).

- The **Genomic View** is the primary display used for interpretation and classification of aberration data. This tab offers an interactive graphic representation of the genome, including probe signal data, detected aberrations, SNP data and annotation tracks. Multiple datasets can be overlaid in this view.
- The **Table View** lists all probes alongside associated data such as start and stop positions, green and red signal intensities and log ratios.
- The **CNVs/SNPs** tab shows all aberrations and their details, either through an ideogram or a table view.

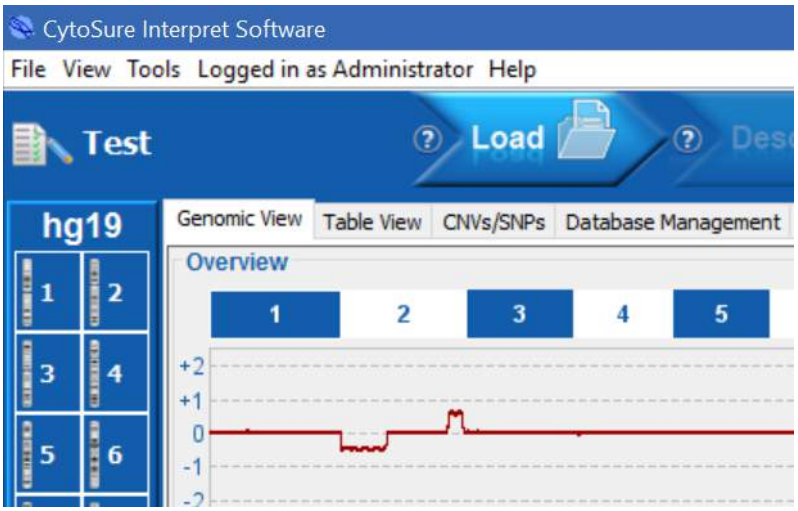


Figure 1: Top left section of the CytoSure Interpret Software screen showing the first four main tabs: Genomic View, Table View, CNVs/SNPs and Database Management.

Genomic View Tab

The **Genomic View** tab is main the screen, where most visual analysis will take place. It consists of four sections, as outlined below. It is possible to orientate this view both Horizontally (the default) and Vertically. To change the orientation, select **View -> Switch Orientation**). In general, this guide describes the appearance and functionality in the software based on horizontal orientation.

Overview section

The **Overview** section (Figure 2) displays all CNV probe and aberration data for either the selected chromosome or the whole genome (as chosen in the **Chromosome Navigation Toolbar** [Figure 5]):

Viewing CNVs

The **Overview** section contains a chromosomal diagram at the top (or, if the graph represents the whole genome, numbered blocks represent chromosomes). The graph directly below this diagram indicates probes and CNVs across the chromosome (or genome). Probes are displayed as circles on the graph, positioned according to the normalised, base-2 log ratio of the signal intensity of the probe in the sample against that of the reference. Detected CNVs are displayed as shaded blocks (or horizontal bar if zoomed out) in the location of detection. The colour of the block represents the colour of the sample. In Figure 2 there are two detected CNVs on 2q. Only one sample is shown, which is blue in colour.

Sample colours are automatically assigned, but can be edited in the **Options...** menu in the **Colour Scheme** section at the bottom of the **Display** tab, where **Colour 1** is associated with the first sample opened and **Colour 2** the second etc. To see which colour has been assigned to a sample, navigate to the **Genomic View** tab and observe the sample tabs at the far right of the page, as seen in Figure 3. The colour of the sample name text is the colour associated with the sample.

Clicking on a chromosome or one of its karyotype bands (in the top diagram) will highlight the area, surrounding it in a dotted box (as seen in Figure 2 near q14.3). Alternatively click and drag on the graph to manually highlight a region. The highlighted region will be displayed in the **Chromosome Section**.

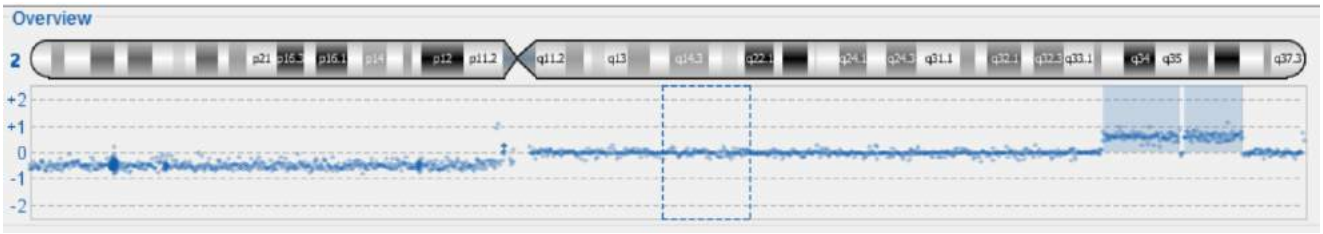


Figure 2: Overview section of Genomic View showing probe data and detected CNVs for chromosome 2. Blue dotted box indicates which chromosome portion is displayed in the Chromosome Section. The shaded blue sections indicate detected CNVs.

Viewing SNPs/LOH

If you right-click on any part of the **Overview Section**, two options will be displayed: **Show B-Allele Frequency Plot** or **Show Allele Status Plot** and **Enable** or **Disable SNP Data Zooming**. The options present depend on which settings are currently active.

Choose the first option (**Show...**) to bring up another plot below the CNV graph (Figure 3). If this plot is a B-allele frequency plot then dots on the graph indicate the position of the SNP (x-axis) and the frequency of the alternate allele (y-axis). If the plot is an allele status plot then the y axis will indicate the genotype of the SNP, as determined using settings which can be edited through the **SNPs/CNVs** tab (in the ideogram view, hover the mouse over the

centre of the bottom of the page and click on the **Options...** button). Switch between the plot types by right-clicking on the area and choosing **Switch to...Plot**.

The second option is used to switch between enabled and disabled zooming on the SNP plot. Zooming means that you can select an area of the chromosome to zoom in on by clicking and dragging on the SNP plot, in the same way as can be done on the CNV plot above.

Detected LOH are shown by a shaded red section (or horizontal bar if zoomed out).

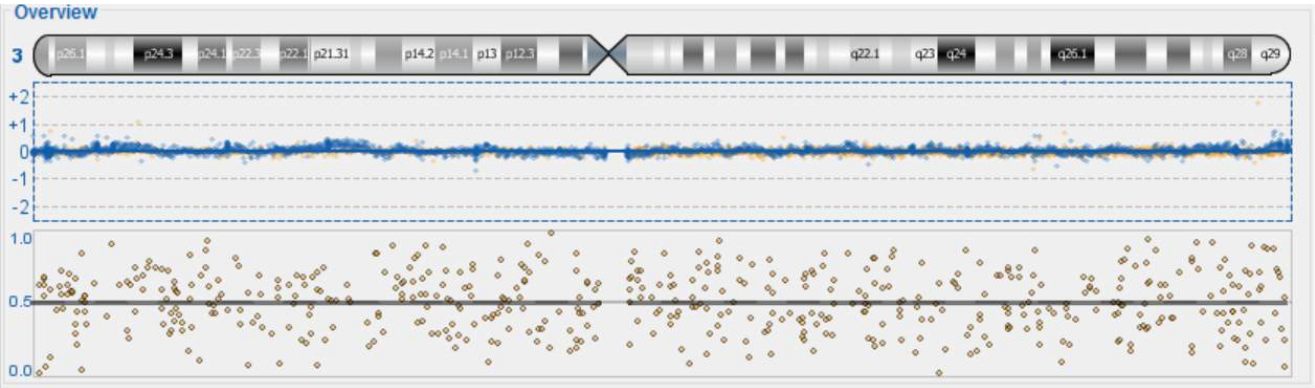


Figure 3: Overview section of Genomic View showing B-allele frequency plot for the 'yellow' sample.

Chromosome Section

The **Chromosome Section** (Figure 4) displays a more detailed version of the **Overview** section graph.

Viewing CNVs

The X-axis displays the location in bases from the start of the chromosome. Probes are displayed as circles on the graph, positioned according to the normalised, base-2 log ratio of the signal intensity of the probe in the sample against that of the reference. Hover over the circle to bring up probe information. Faint grey vertical lines indicate the edge of karyotype bands, which are labelled in grey on the X-axis (at the top) and finally detected CNVs are displayed as shaded bars, as seen in blue in Figure 4.

> If you're graph does not look quite like Figure 4 then don't panic! There are numerous display settings which alter how it looks. These can be accessed by choosing **Tools -> Options... ->Display**.

There are numerous ways to select or change the location displayed in this section:

- Highlight a region in the **Overview** section**.**
- Manually highlight a region in the **Chromosome Section** by clicking and dragging the dotted box.
- Use the zoom in and out buttons at the top left of the right side panel (Figure 5).
- Use a mouse scroll button or left-click to move the window up or down the chromosome.
- Click on a chromosome in the **Chromosome Navigation Toolbar** to view a different chromosome (or the whole genome).
- Click on an aberration in the **Aberration Table** (in the right side panel) to display the aberration.

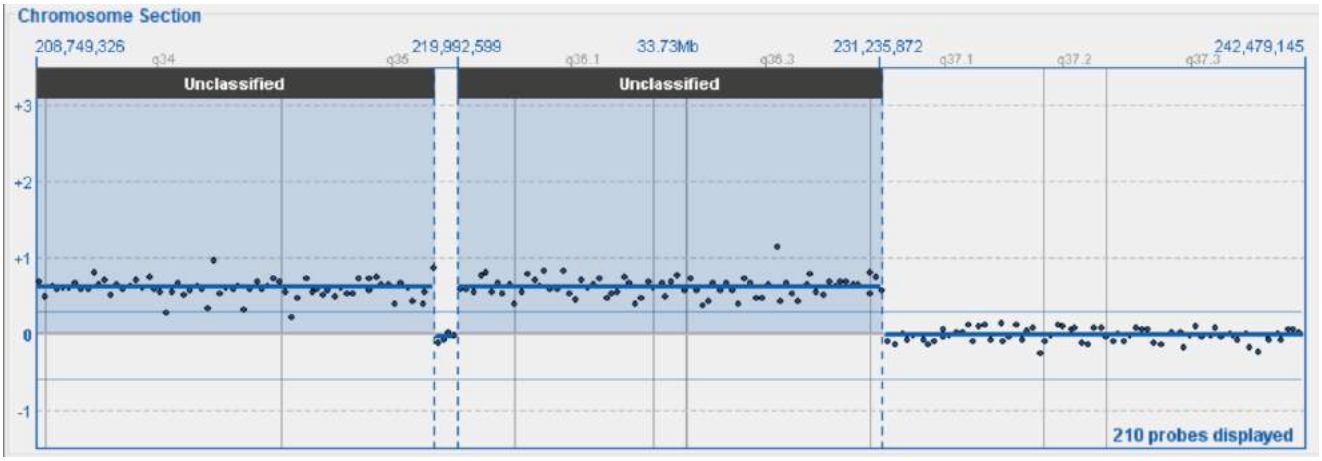


Figure 4: Chromosome Section of Genomic View showing probe data and detected CNVs for a portion of chromosome 2.

Figure 5: Chromosome Section of Genomic View showing a detected CNVs and the CNV right-click menu options.

Chromosome Navigation Toolbar

The **Chromosome Navigation Toolbar** is used to switch between a chromosome or whole genome view in the **Chromosome** and **Overview sections**. The white highlighted box indicates what is currently selected (Figure 6).

Figure 6: Chromosome Navigation Toolbar of Genomic View. Chromosome 3 is selected.

Right-hand panel

See Figure 7 for an example panel. Starting from the top of the panel and working downwards, the displays and options are as follows:

- **Zoom in and out buttons**:
    - Found in the top left corner of the panel, these are used to alter the display window of the **Chromosome Section**.

- **Other options in the blue section** (at the top of the panel)**:**

  These are used to specify what to display in the Overview and Chromosome Sections.

- The **Raw** option displays raw (identical to the values imported from the original feature extracted file) probe data only.
- The **Normalised** option displays normalised probe data with detected CNVs.
- The **Display Filtered?** option will display filtered out probes in grey on top of the black non-filtered probes.
- The filtered icon () is greyed out when no probe filters are applied. Alternatively, when the icon is white, filters have been applied and clicking on the icon brings up a **Probe filtration status** window, displaying the filters used.

> Edit probe filters by choosing **Tools -> Options... -> Probe Filtering**.

- **Vertical tabs:** The right-hand side (of the panel):
  - These tabs, which are found on the right-hand side of the panel, are associated with each open dataset (1 tab per dataset/sample). In Figure 7, only one dataset is open.
  - The colour of the sample name text on each tab indicates the colour associated with the sample. These colours are used in several different places, such as the probes and CNV colours in the graphs on this page. The colours associated with each sample can be changed by navigating to the **Colour Scheme** section at the bottom of the Display tab in the **Options...** menu, where **Colour 1** is associated with the first sample opened and **Colour 2** the second etc..
  - The tab buttons are as follows:
    - Top tick box () — single click to show/hide the dataset. Double-click to show only this dataset.
    - Top save button () — Saves changes to the sample to file.
    - Bottom save button () — Saves changes to the sample to the database.
    - Bottom tick box () — Inverts the data (Cy5/Cy3 and vice versa).
    - Close button () — Closes the dataset.
- **Sample Information box:**
  - This box displays sample information and metrics.
  - Clicking **Edit** will open the **Sample Details** window, from which sample and experiment data can be altered.
  - Clicking **Customise** will open a window from which to specify which sample and experiment details are displayed.

- **Aberrations Table:**
  - Found at the bottom of the panel, this table lists CNV and LOH regions.
  - Click the table heading to sort the entries by that category.
  - To display a CNV in the **Chromosome Section**, either left-click on the associated row in the table or right-click and choose **View Aberration** from the options menu. Bold table entries are those which are displayed in the **Chromosome Section**.
  - To bring up the **Aberration Details** window for a CNV, either double left-click on the associated row in the table or right-click and choose **Edit Aberration Details** from the options menu. Some details can be edited in this window.
  - To view an aberration **Signal Intensity Comparison** graph for a CNV, right-click on the associated row in the table and choose **View Signal Intensity Comparison** from the options menu.

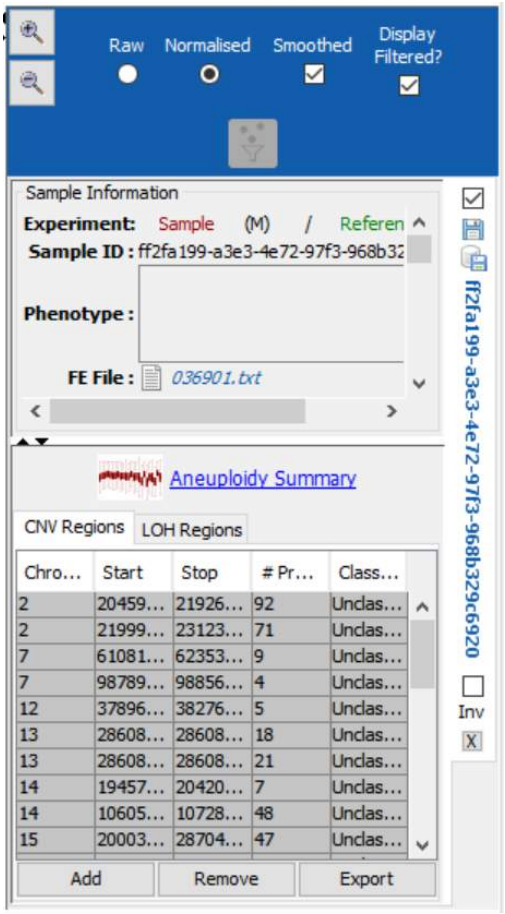
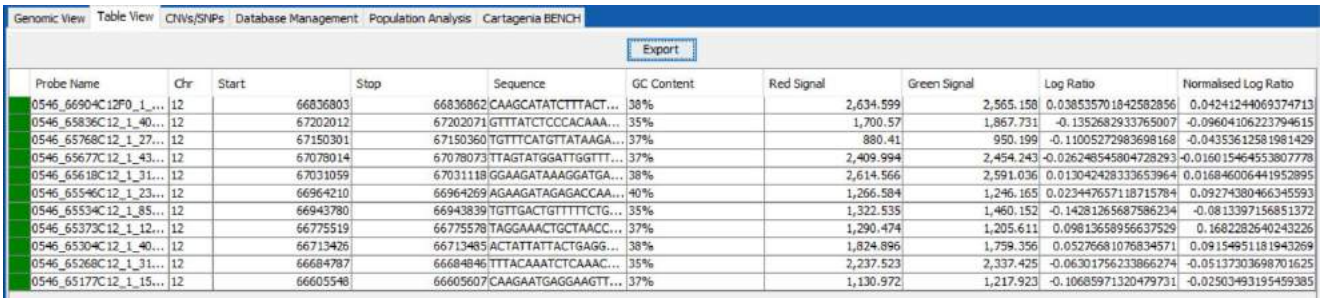


Figure 7: Right Side Panel of Genomic View.

Table View tab

The **Table View** tab is an alternative way in which to view the data. This display contains a table of all CNV probes and the details associated with them: probe name, chromosome, start and stop positions, sequence, GC content, red and green signal, log ratio and normalised log ratio. An example of the **Table View** tab can be seen in Figure 8. The coloured boxes in the left-most column (green in Figure 8) indicate the sample from the which the probe information comes from. In Figure 8, all displayed probes are from one sample. Sample colours are automatically assigned, but can be edited in the **Options...** menu in the **Colour Scheme** section at the bottom of the Display tab, where **Colour 1** is associated with the first sample opened and **Colour 2** the second etc.. To see which colour has been assigned to a sample, navigate to the **Genomic View** tab and observe the sample tabs at the far right of the page, as seen in Figure 9. The colour of the sample name text is the colour associated with the sample.



Probe Name	Chr	Start	Stop	Sequence	GC Content	Red Signal	Green Signal	Log Ratio	Normalised Log Ratio
0546_66904C12F0_1_...	12	66836803	66836862	CAAGCATATCTTTACT...	38%	2,634.599	2,565.158	0.038535701842582856	0.042412440693747 13
0546_65836C12_1_40...	12	67202012	67202071	GTTTATCTCCCACAAA...	35%	1,700.57	1,867.731	-0.13526829337650 07	-0.09604106223794615
0546_65768C12_1_27...	12	67150301	67150360	TGTTTCATGTTATAAGA...	37%	880.41	950.199	-0.11005272983698168	-0.04353612581981429
0546_65677C12_1_43...	12	67078014	67078073	TTAGTATGGATTGGTTT...	37%	2,409.994	2,454.243	-0.026248545804728293	-0.016015464553807778
0546_65618C12_1_31...	12	67031059	67031118	GGAAGATAAAGGATGA...	38%	2,614.566	2,591.036	0.013042428333653964	0.016846006441952895
0546_65546C12_1_23...	12	66964210	66964269	AGAAGATAGAGACCAA...	40%	1,266.584	1,246.165	0.023447657118715784	0.09274380466345593
0546_65534C12_1_85...	12	66943780	66943839	TGTTGACTGTTTTTCTG...	35%	1,322.535	1,460.152	-0.14281265687586234	-0.08133971568513 72
0546_65373C12_1_12...	12	66775519	66775578	TAGGAAACTGCTAACC...	37%	1,290.474	1,205.611	0.09813658956637529	0.16822826402432 26
0546_65304C12_1_40...	12	66713426	66713485	ACTATTATTACTGAGG...	38%	1,824.896	1,759.356	0.052766810768345 71	0.09154951181943269
0546_65268C12_1_31...	12	66684787	66684846	TTTACAAATCTCAAAC...	35%	2,237.523	2,337.425	-0.06301756233866274	-0.05137303698701625
0546_65177C12_1_15...	12	66605548	66605607	CAAGAATGAGGAAGTT...	37%	1,130.972	1,217.923	-0.10685971320479731	-0.02503493195459385

Figure 8: Table View tab showing list of probes (and their associated information) for chromosome 12.

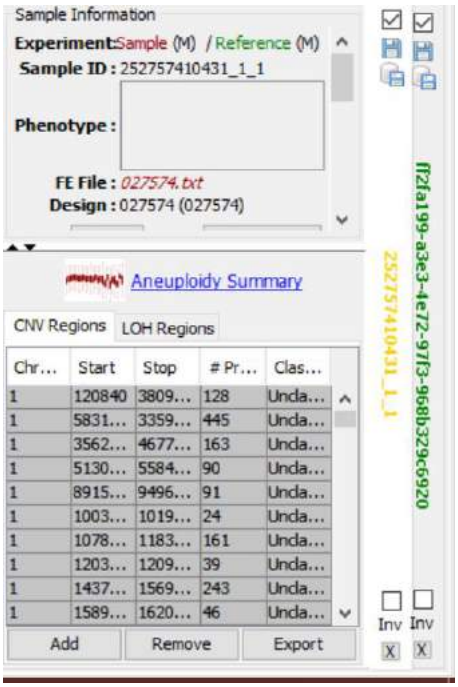


Figure 9: Right hand side of Genomic View tab showing the sample tabs of two samples; one green and one yellow.

Personalise view

The columns displayed in the table can be altered by right-clicking on the table header. This will bring up a list of probe details types, where a blue tick to the left of the text indicates that the field is displayed in the table and a blank box indicates that it is hidden (see Figure 10). Choose which fields to display by clicking on the row in the menu to select or deselect it. Additionally, the order of the columns can be altered by left-clicking on the header and dragging it horizontally to the required location. Rows in the table can be sorted by each header by left-clicking on the header: click once to order in one direction and twice to order in the opposite direction. A small arrow should appear in the header box indicating the direction of sorting — increasing or decreasing.

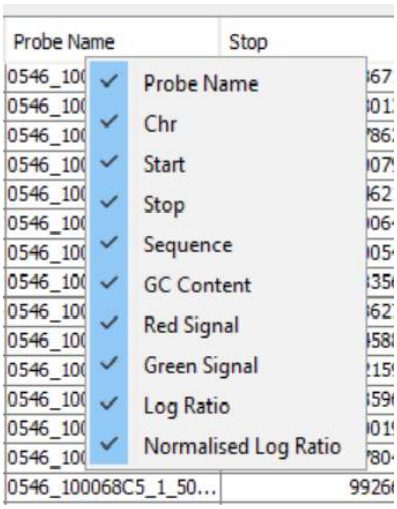


Figure 10: Table View header right-click menu.

Export data

There are two ways to export the data presented in this table:

1. Copy and paste — It is possible to highlight subsections of the data by clicking and dragging. Press *Ctrl-c* on the keyboard to copy the selected data.
2. Export function — There is an **Export** button in the centre directly above the table. Clicking this button will save a tab-delimited version of the table in a text file in the location chosen in the save window. Row order is not preserved and only the columns displayed will be exported. All rows will be saved.

CNVs/SNPs tab

The **CNVs/SNPs** tab (Figure 11) offers two alternative ways in which to view aberrations; as a table or an ideogram. As with the **Genomic View** tab, the **CNVs/SNPs** tab can be split into several sections, as outlined below.

All **.cgh** files which are currently opened in the software will be displayed in the **CNVs/SNPs** page. Use the tabs labelled with sample IDs to switch between sample views.

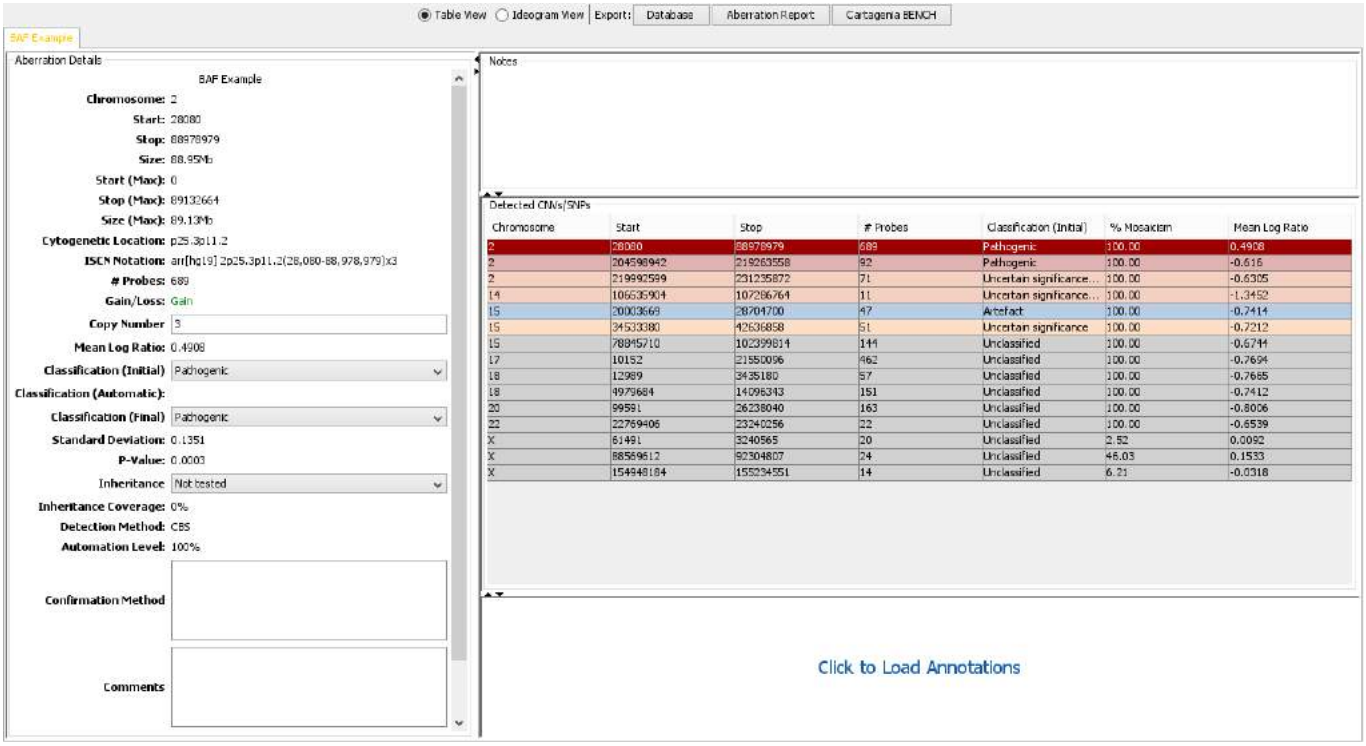


Figure 11: CNVs/SNPs tab displaying the table view.

Top buttons section

These buttons encompass two major functions: to alter the data view and to export the data.

- **Table View** and **Ideogram View** radio buttons:
    - These alter what is displayed in the **Detected CNVs/SNPs** section of the page (Figure 11, middle right).
- **Database** button:
    - Used to export the aberration details to the database.
- **Aberration Report** button:
    - This will bring up a window asking whether you wish to export to a PDF, a HTML or a tab-delimited text file.
    - Each option will create a file, saved in a location of your choice, which lists all detected aberrations and their classifications, as well as overview information on the sample, analysis and QC results.

- Additionally, the PDF and HTML reports will display the ideogram view of aberration, as well as additional aberration display sections for each aberration (Figure 12).
- Please note that this includes both CNVs an LOH regions.

- **Cartagenia BENCH** button:
  - Uploads aberrations to Cartagenia BENCH, which can be accessed through the **Cartagenia BENCH** tab.
- **Show CNVs?** and **Show SNPs?** tick boxes:
  - These options are only visible when the ideogram view is active.
  - They determine whether CNVs and/or SNPs are displayed in the **Detected CNVs and SNPs** section of the page.

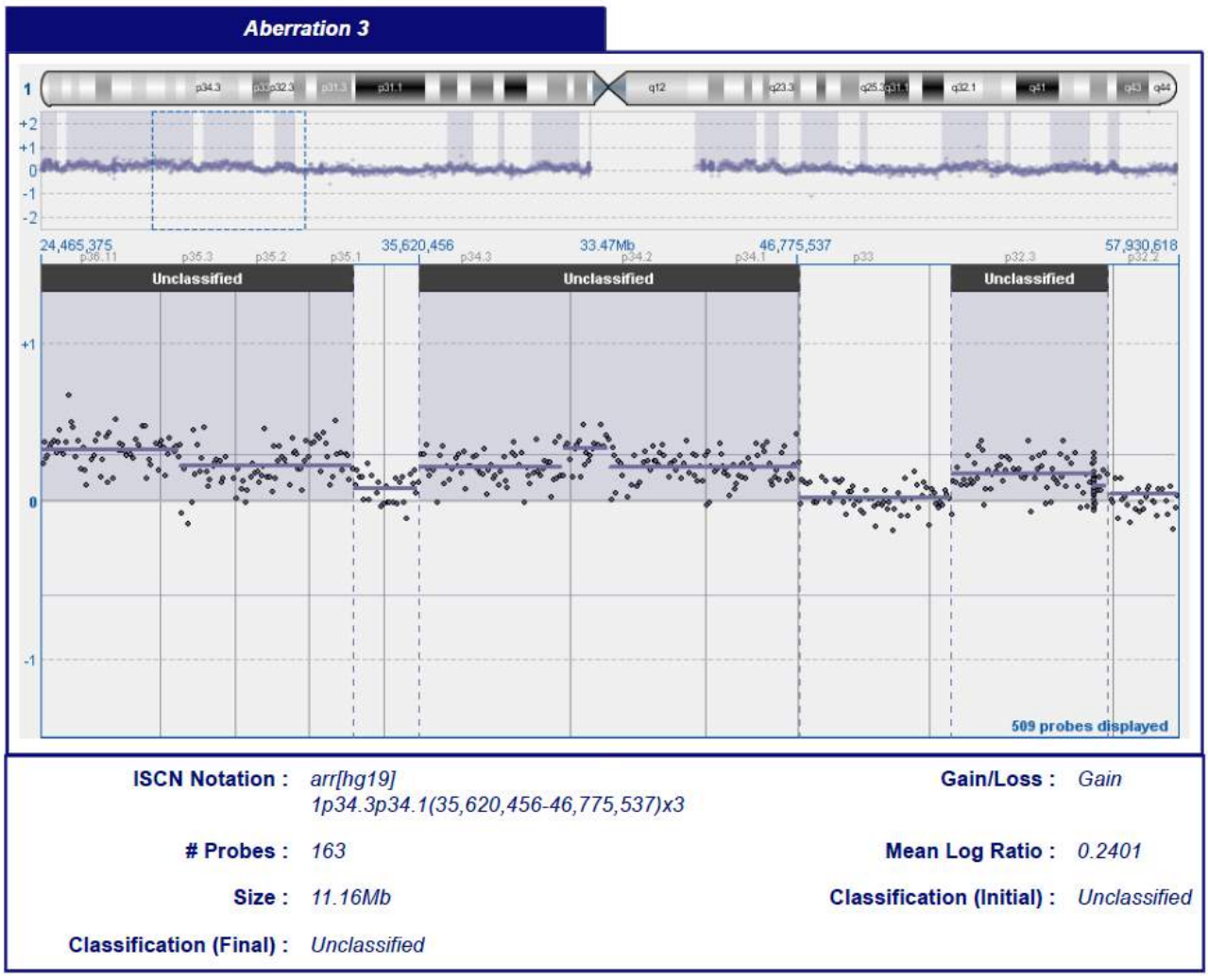


Figure 12: Example Visual Aberration Display included in HTML and PDF aberration reports.

The pairs of small black triangles at the edge of each section of the **CNVs/SNPs** page are used to minimise or maximise the page section. If a section appears to be missing on your view, try clicking the little black arrows, or dragging the other section edges to uncover the missing section.

The Aberration Details section

This section of the page, which displays additional aberration details, can be seen on the left-hand side of Figure 11. The details loaded will correspond to the aberration highlighted/selected in the **Detected CNVs/SNPs** section. Some of these details, such as classification and inheritance, can be altered here.

The notes section

As seen near the middle right of Figure 11, this section displays sample associated notes. It is possible to edit these notes here.

The detected CNVs/SNPs section

This section of the page, which displays all aberrations detected in the sample, can be seen near the middle right hand side of Figure 16. The view displayed here depends on which view is selected, **Table View** or **Ideogram View.**

Table view

CNVs are listed, alongside basic information such as start and stop sites, number of probes and classification. Aberrations of different classification will be displayed in different colours. Double-click on an aberration row in the table to bring up a separate window of additional details (which can also be seen to the right in the **Aberration Details** section). Right-click on an aberration row to bring up additional options, such as to choose a classification.

To view or edit the colours assigned to classifications, navigate to **Tools -> Options... -> Classification.**

Ideogram view

Each chromosome is displayed visually with CNVs and/or SNPs indicated at either side. If the displayed data includes SNPs then the two tick boxes, **Show CNVs?** and **Show SNPs?**, under the export options near the top of the page, can be used to indicated whether to show CNVs, SNPs or both are displayed.

- **CNVs** — If only CNVs are displayed, the chromosomes should resemble Figure 13. Green rectangles indicate gain of CNVs and red ones loses of CNVs. Please note that long rectangles may represent more than one CNV. Click on the rectangle to select it.
- **SNPs** — If only SNPs are displayed, the chromosomes should resemble Figure 14. Black bars to the right side of the chromosomes represent heterogyzous SNPs. Red bars to the left indicated homozygous SNPs. If the user clicks or hovers over the bars, they will become bold and rounded. The percentage represents the proportion of homozygous SNPs.

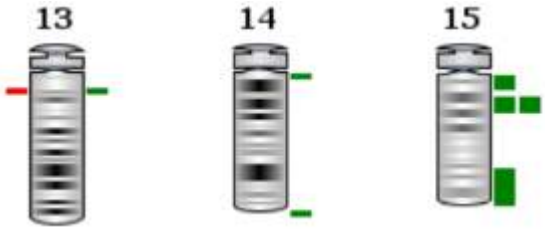


Figure 13: Ideogram view of chromosomes 13-15 displaying detected CNVs. Green rectangles indicate gain CNVs and red ones loses.

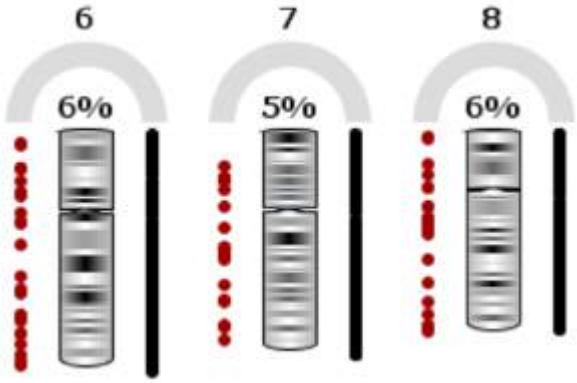


Figure 14: Ideogram view of chromosomes 5-7 displaying detected SNPs. Black bars represent heterogyzous SNPs, red bars homozygous ones.

Use the **Options** button at the middle of the bottom of the page (you may have to hover over it to make it appear) to bring up an options menu (Figure 15). From this, you can edit a number of SNP settings including:

- Limits for genotype assignment.
- Homozygosity threshold.

- Application of probe filters.
- LOH calling region size.
- LOH reporting.
- Assigned colours of homozygous and heterozygous SNPs (by default these are red and black respectively). This will affect the B-Allele frequency plots, the distribution plot and the ideogram view of the SNPs.

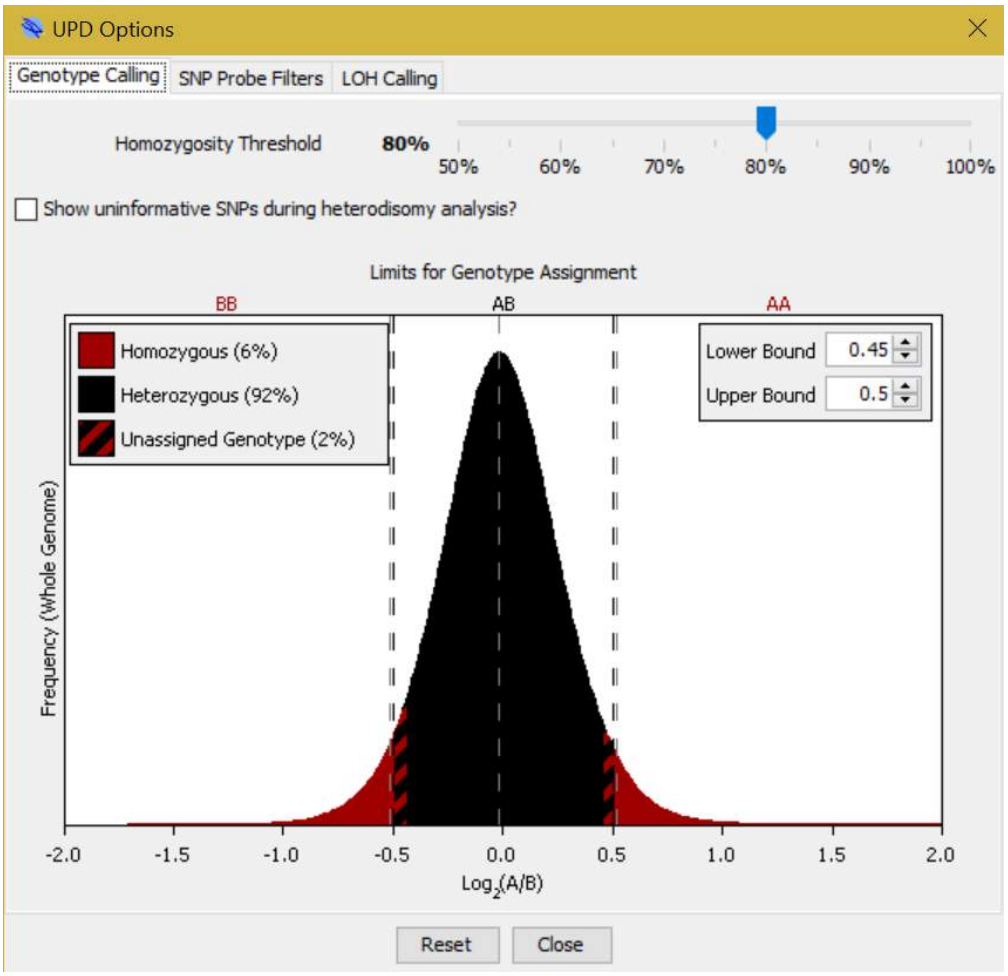


Figure 15: UPD Options menu.

Click on a chromosome image to view it in more detail. This will load the **Detected CNVs/SNPs** detailed chromosome view (Figure 16).

- The original chromosome diagram with CNV and SNP bars is shown along the centre horizontal section of the view.

- Above is a small graph of percentage homozygosity against number of SNPs for all chromosomes (with the currently viewed chromosome in red). Additionally, if LOH regions have been detected, they will be displayed in a table. The red x button is used to return to the non-detailed view.
- Below the chromosome diagram is a plot — either a B-allele frequency plot or a distribution plot. Right-click and choose the relevant option to switch between these views.
- The options menu can also be accessed from here using the **Options** button which appears when you hover over the centre bottom of the page.

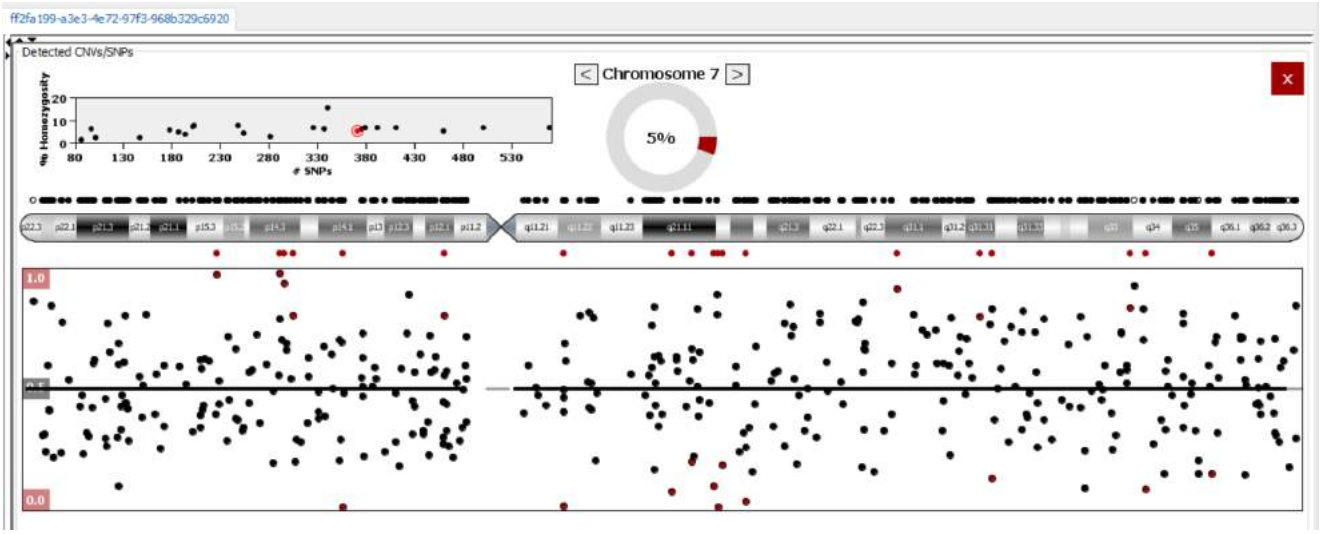


Figure 16: Detected CNVs/SNPs Detailed Chromosome View with B-allele frequency plot.

View annotation tracks

The **Annotation Track** panel is located at the bottom of the **Chromosome Section** of the **Genomic View** tab. In this section, a number of different annotation tracks can be displayed to aid interpretation of the data. Figure 17 shows an example set of annotation tracks.

Note: If no tracks are displayed, the **Annotation Track** panel will not be visible.

The information in the tracks (apart from the LOH and aberration tracks) is obtained from various external sources, stored in the CytoSure Interpret Software database and can be

updated following updating to a new software release. Additionally, users may create custom annotation tracks from their own data or other external resources if required (via **Tools -> Options... -> Annotation -> Import new track**). The source of built-in annotation tracks can be identified by choosing the **Annotation** tab from the **Options...** menu, double-clicking on the annotation of interest in the table and observing either the **URL Composition** or the **File Location** fields. The **LOH** and **Aberration** tracks show LOH regions and CNVs identified in all samples previously submitted to the database.

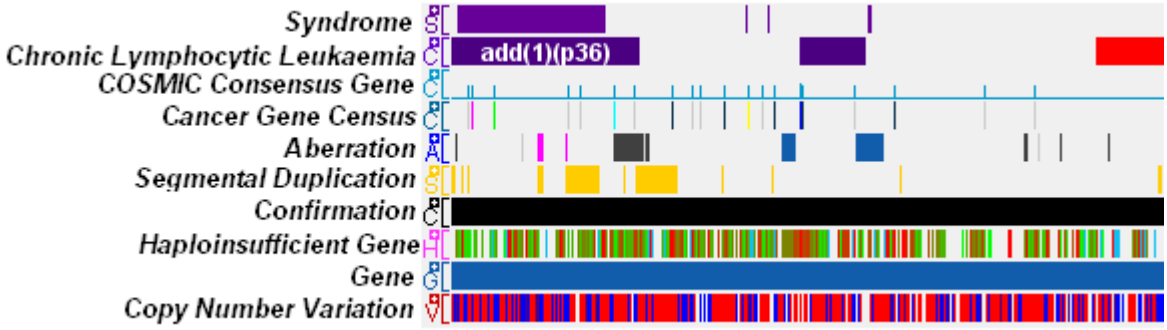


Figure 17: Example Annotation Track panel.

Displaying annotation tracks

To display annotation tracks, right-click on the **Chromosome Section** in the **Genomic View** tab. This will bring up an options menu, as seen in Figure 18. All available tracks will be listed in this menu. The text colour represents the colour the track will have when loaded. It is also possible to import additional tracks by choosing **Tools -> Options... -> Annotation -> Import new track**.

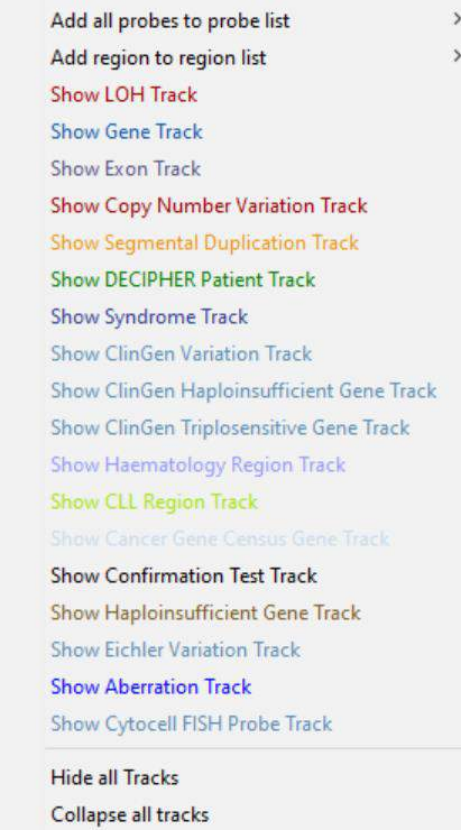


Figure 18: Options menu for annotation tracks.

Once a track is selected, it will appear as a bar at the bottom of the **Chromosome Section** (Figure 19). The letter at the right-hand side of the bar (which is normally the first letter of the track name) is the identifier for the track. Hover over the identifier to display the full track name. The bars displayed in the track represent the location of annotations/features relative to the **Chromosome Section** graph.

There are three ways that an individual annotation might be coloured:

- If the individual annotation was assigned a colour code when it was imported into the database, it will be displayed in that colour.
- Otherwise, it will be displayed in the same colour as has been assigned to the track.
- Additionally, if both **Enable On-Hover Annotation Colour Changes** and **Invert On-Hover Annotation Colour Changes** in **Tools -> Options… -> Display** are selected, and the annotation represents either a gain or a loss (e.g. if it is part of an external CNV database) then it will be coloured according to its gain/loss status (i.e. Red for loss, Blue/Green for gain).

Figure 19: Chromosome Section with three loaded annotation tracks: Gene, Segmental Duplication and CLL.

There are two primary display types for the annotations: annotation details (as seen in Figure 19) and frequency plot. Switch between the two either by left-clicking once on a blank space in the track, or by right-clicking on a blank space and selecting **Show Frequency Plot** or **Show Annotation Details** from the menu. When the annotation details are displayed, hovering over an annotation will display additional details. See the Types of information displayed section for more details.

Edit track display

There are numerous ways in which to alter the display of the annotation tracks:

Display order

Right-click on a track in the **Annotation Panel** (not on top of an annotation) to bring up a menu. Choose **Move Up** or **Move Down** to move the track higher or lower in the panel.

Type of information displayed

The annotation bars displayed in the track can represent either a **Frequency Plot** or **Annotation Details**.

- In the **Frequency Plot** (see Figure 20 for an example), the horizontal position of the bars indicate the location of the feature and the vertical height the frequency (the amount of overlapping features or the frequency of occurrence).
- In the **Annotation Details** display, the horizontal position of the bars still indicates the location of the feature, but they are all the same height. When in the expanded view of the track, multiple features at the same position are displayed on top of each other, but in the collapsed view they are combined together. When zoomed in far enough, the annotation label will display on the bar. When the mouse is hovered over the bar,

additional annotation details will be displayed and right-clicking on the bar will bring up an additional menu with options to link out to external data resources.

Switch between the two by either left-clicking once on a blank space in the track, or right-clicking on a black space and selecting **Show Frequency Plot** or **Show Annotation Details** from the menu.

Figure 20: Section of Gene and Segmental Duplication annotation tracks in frequency plot view.

Size

Individual annotation tracks can either be **Collapsed** or **Expanded**:

- **Collapsed** tracks use one row to display all features in the track, which may result in annotations overlapping each other. The **Frequency Plot** option of the collapsed view will show a graph in which the height represents the number of overlapping features at that position, but with total height limited to one row.
- **Expanded** tracks may use multiple rows to display their features, the number of rows being determined by the **Maximum Track Width** option (see the Other options section). In the **Annotation Details** view, each of the features present will be displayed in a similar way to in the collapsed view, but each overlapping feature will be placed on top of the other, as seen in Figure 21. In the **Frequency Plot** view, the plot will be the same as in the **Collapsed** view but stretched to cover all rows.

Figure 21: Annotation Details Expanded view of Gene annotation track showing ST8SIA6 gene and two additional overlapping genes (ST8SIA6-AS1 and Y-RNA).

To switch between the **Collapsed** and **Expanded** views:

- Right-click on a track in the **Annotation Panel** (not on top of an annotation) to bring up a menu. Choose **Expand** or **Collapse** to alter the height of the track.
- Right-click on the **Chromosome Section** and choose **Expand/Collapse all tracks** from the menu. From this menu, it is also possible to hide the annotation panel by choosing **Hide all tracks.**
- Locate the small square above the annotation track initial on the left side of the track (as seen in Figure 22) containing either a "+" or a "-". Click this to either expand or collapse the track.

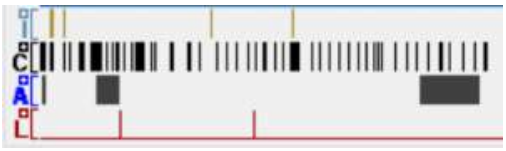


Figure 22: Four annotation tracks in the Collapsed view.

Other options

A number of other options can be edited from the **Options...** menu. To do this, choose **Tools -> Options... -> Annotation**. This will open the screen shown in Figure 23. Click on the track of interest in the top table to highlight it. Then choose the **Edit selected** button to open an additional window (Figure 24). From this window, it is possible to edit the **Track Name, Identifying Initial, Colour, URL Composition, Maximum Track Width** and **Maximum # Displayed.**

The **Maximum Track Width** is the maximum height of the expanded view, where 1 represents the height of the collapsed view. **Maximum # Displayed** is the maximum number of features displayed on one screen. If there are more features present than permitted, the track will display **Zoom In**. This can be used to avoid waiting for high density track information at a resolution that is too small to be informative.

URL Composition

This is the composition of the URL which will be displayed when this link is clicked. A number of variables can be included in this field to enable the software to dynamically generate URLs based on information relating to the selected annotation track (e.g. annotation name, chromosome, start position, stop position etc.). For example, to create a link that searches the OMIM database for entries that match the name of the selected annotation, the URL could be set to “http://omim.org/search?search={#name}”. The text “{#name}” will be replaced with the name of the selected annotation when the user clicks on the link to create, for example, “http://omim.org/search?search=DMD”. Available variables include:
#name — Name of the annotation.

- #chrName — Chromosome name (e.g. X)
- #chrInt — Chromosome index (e.g. 23)
- #start — Base-pair minimum span start position
- #stop — Base-pair minimum span stop position
- #maxStart — Base-pair maximum span start position

- #maxStop —Base-pair maximum span stop position
- #build — Genome build (e.g. hg18)

Users can also include the values of extra annotation information (i.e. Columns labelled as “Other” during annotation import) in URLs. An example of this is the “PubMed” annotation link for the Copy Number Variation track, whose link is “http://www.ncbi.nlm.nih.gov/pubmed/{PubMedID}”. The same technique can be used to complete the “URL" field of an annotation link.

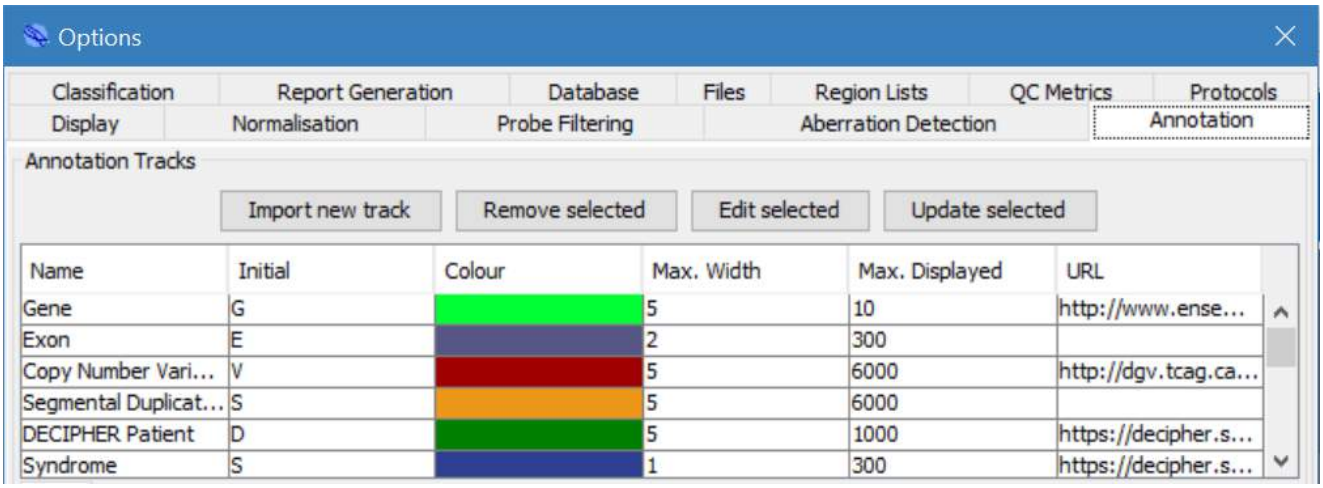


Figure 23: Section of Gene and Segmental Duplication annotation tracks in Frequency Plot view.

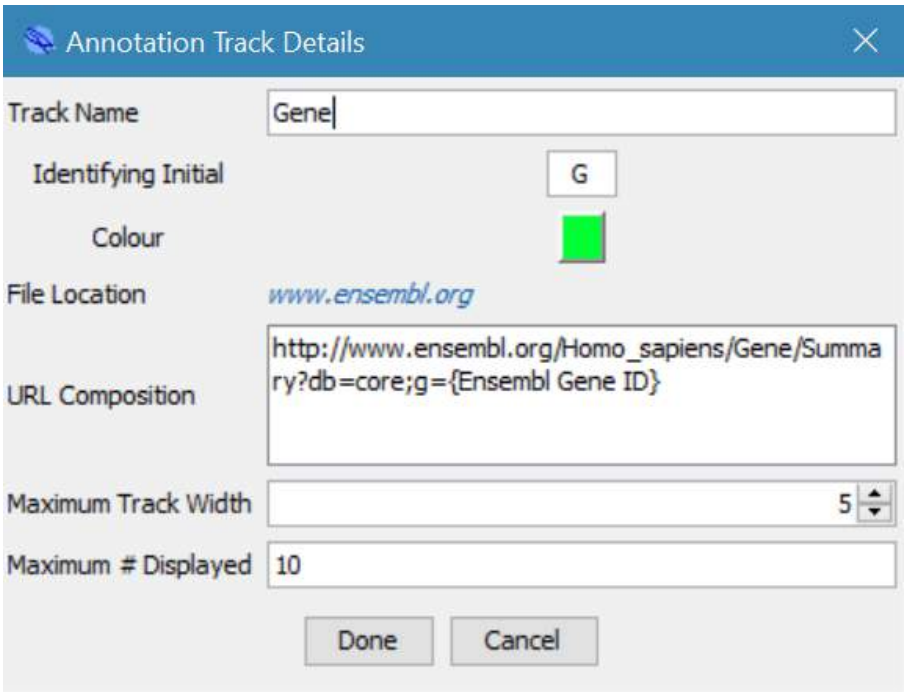


Figure 24: Annotation Track Details window used for editing annotation tracks.

External links

In order to increase the quantity of data available for analysis, the annotation tracks contain a number of options to link out to external data sources.

Opening links

Right-click on a feature in an annotation track to bring up the **View In** menu, which contains a number of options. The options on the menu vary depending on the annotation track, but by default should contain the **View Region in...** options, as seen in Figure 25. Clicking on each of these will open a webpage (provided you have internet access) of the selected region in the selected online resource. Some of the annotation tracks have more options, such as **View Gene In...** options from the **Gene** annotation track and **View Copy Number Variation In...** from the CNV annotation track.

Figure 25: Most simple form of View In menu, accessed by right-clicking on a feature in an annotation track.

Additionally, for some of the tracks, one left-click on a feature will open an external webpage in a browser. For example, by default, clicking on a gene in the **Gene** annotation track will open the relevant gene summary page in Ensembl.

Editing links

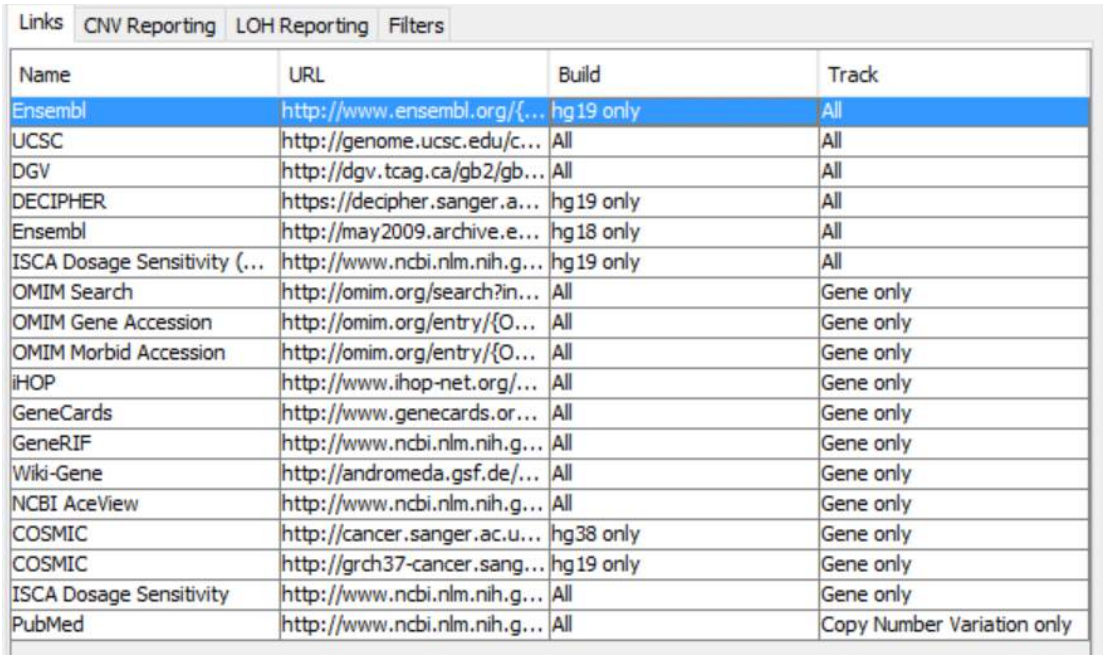
Users can customise the links available in each annotation track via **Tools -> Options... -> Annotation -> Links**. An example of this section can be seen in Figure 26. To edit details of an existing link, left-click on the entry in the table to highlight it and then choose the **Edit** button to open the **Edit Annotation Link** window. From this window, it is possible to edit the **Name, URL, Build** and **Track** option. The URL Composition describes the variables that can be used in the URL field to generate dynamic links based on the selected annotation.

Additionally, different URLs for individual annotations may be specified in each row of the text file describing the annotations in the track.

Creating links

Additionally, there are two ways to create new links:

1. Click the **Duplicate** button at the bottom of the **Options... -> Annotation -> Links** page to generate a duplicate track. Edit the details as necessary and save.
2. Click the **Create** button at the bottom of the **Options... -> Annotation -> Links** page. Enter details of the new track into the window and save.


Links | CNV Reporting | LOH Reporting | Filters

Name	URL	Build	Track
Ensembl	http://www.ensembl.org/{...	hg19 only	All
UCSC	http://genome.ucsc.edu/c...	All	All
DGV	http://dgv.tcag.ca/gb2/gb...	All	All
DECIPHER	https://decipher.sanger.a...	hg19 only	All
Ensembl	http://may2009.archive.e...	hg18 only	All
ISCA Dosage Sensitivity (...	http://www.ncbi.nlm.nih.g...	hg19 only	All
OMIM Search	http://omim.org/search?in...	All	Gene only
OMIM Gene Accession	http://omim.org/entry/{O...	All	Gene only
OMIM Morbid Accession	http://omim.org/entry/{O...	All	Gene only
iHOP	http://www.ihop-net.org/...	All	Gene only
GeneCards	http://www.genecards.or...	All	Gene only
GeneRIF	http://www.ncbi.nlm.nih.g...	All	Gene only
Wiki-Gene	http://andromeda.gsf.de/...	All	Gene only
NCBI AceView	http://www.ncbi.nlm.nih.g...	All	Gene only
COSMIC	http://cancer.sanger.ac.u...	hg38 only	Gene only
COSMIC	http://grch37-cancer.sang...	hg19 only	Gene only
ISCA Dosage Sensitivity	http://www.ncbi.nlm.nih.g...	All	Gene only
PubMed	http://www.ncbi.nlm.nih.g...	All	Copy Number Variation only

Figure 26: Links table in Annotations tab of Options... menu.

Quality metric review

One of the most important aspects of any microarray experiment is data quality. Before carrying out any in-depth analysis, it is vital to ensure that the data you are working with has not been comprised by any technical issues. You need to be sure that the aberrations you are detecting are caused by biology not by how the experiment was performed.

Using CytoSure Interpret Software, it is very easy to view the QC metrics and to keep track of how they vary over a particular experiment.

Types of quality metric

Many factors can influence the expected range of the reported quality metrics, including the source and quality of the DNA, the protocols used, the scanner and image processing software. Below is an explanation of each metric provided by the software. Table 1 displays the default parameters for good quality data: these are guidelines that have been shown to work well using the standard OGT and Agilent protocols.

Please be aware that some of these metrics are specific to SNP calls and will not be reported for arrays without SNP probes.

X-separation

The X-separation is the mean normalised log ratio of the probes on the X chromosome. When using sex mis-matched test and reference sample, the log ratio should be close to 1. For a sex matched test and reference samples, the ratio should be close to 0. There are no quality thresholds for this statistic.

Derivative log ratio (DLR) spread

This is perhaps the most important QC metric. It calculates the probe-to-probe log ratio noise of an array. A poor DLR spread will mean that it is more difficult to accurately call amplification or deletion calls.

Background noise

This metric is calculated as the standard deviation of negative control probes on the array. Initially any outlier spots are rejected and the standard deviation of the remaining spots is calculated. If this value is high, the tiff image should be examined. If areas of high background noise are evident, the array processing procedure should be investigated. High background values are typically caused during the washing steps. During the washing steps, ensure the stirrer speed is correct — a vortex should be visible when the dish is empty, and ensure the dishes are always properly cleaned. If high background values persist, the dishes should be washed with acetonitrile to remove build-up of un-incorporated dye molecules. A poor background does not necessarily indicate that the array has failed. This is a secondary metric as it is incorporated into the signal-to-noise metric.

Signal intensity

This value is the average raw signal intensity for each channel. There is no upper limit on this value; however, for a CGH array it is very unusual to see saturated spots. This metric is highly dependent on the labelling method used. Therefore, the best way to use this value is to compare it across a set of experiments where the same labelling method has been used. This allows identification of any samples with an unusually low value. Low signal intensity values can be caused by a problem occurring during the setup of a labelling reaction, poor DNA quality or the incorrect amount of DNA being added. These problems should really be identified before the array is set up by using a Nanodrop™ to accurately measure the concentration of DNA and the level of incorporation of dye molecules. Most labelling kits give expected values for the amount of dye incorporation and the concentration of DNA. If your sample meets these recommendations but the signal intensities are poor, a problem with the

clean-up step is likely. This is a secondary metric as it is incorporated into the signal-to-noise metric. A separate measure of signal intensity for SNPs only is provided when appropriate.

Signal reproducibility

This metric is used to measure technical variability across the array. Likely causes of failure are poor mixing during hybridisation caused by:

- Oven failure
- The bubble formed on setting up the hybridisation chamber failing to move freely
- Hybridisation solution leaking from the gasket slide.

Signal-to-noise ratio

This value is calculated by dividing the signal intensity by the background noise and indicates how clearly the spots can be detected above the background level. This metric is dependent on how well the sample labelling and washing steps worked. It is often easier to look at this metric first and then, if it does not pass, identify where the problem occurred by looking at the background noise and the signal intensity. It is difficult to reliably detect aberrations on arrays where the signal-to-noise is low.

Grid alignment

This metric can only be calculated by the feature extraction software, and will either result in a Pass or a Fail. Failure indicates that the grid may not have aligned correctly during the feature extraction process and it may be necessary to manually re-align and re-feature extract the array.

Homozygosity

The total percentage homozygosity of each sample containing SNP data based on the method described by Sund *et al* (Genet Med 2013:15[1]:70–78).

Negative controls

The negative controls on the array are used to measure the non-specific hybridisation that can occur during the processing steps. A high negative control value can indicate that, at some stage during the process, a buffer has become contaminated. Buffer contamination most frequently occurs during the washing step.

Non-uniform features

This is an important metric as it looks at spot quality. The pixels within a good quality spot will have a normal distribution. If, for example, there is contamination by dust or there is a

small scratch over a spot, the normal distribution will be skewed. When a spot is flagged as being an outlier, it will not be removed from the results but it can be filtered out using the filtering tools so that it does not participate in the detection of aberrations. If over 1% of the total number of features are flagged as outliers, the image should be examined to identify the cause of the problem.

Saturated features

A feature is flagged as being saturated if more than 50% of the total number of pixels within a spot are saturated. It is unusual for more than 0.1% of spots on a CGH array to be identified as saturated. If significant saturation is apparent, it is most likely caused by a technical problem, either with too much DNA being added to the hybridisation or a problem with the processing steps. If this occurs, the image would need to be looked at and the labelling metrics examined.

Signal intensity ratio

This value is the ratio between the red and green signal intensities and is calculated from the processed red and green signals from the feature extracted text file. This data has been adjusted for background, with outliers removed and corrected for dye bias. Ideally this value should be close to 1.0. Deviation from this value could indicate a problem with the normalisation method applied during feature extraction. Alternatively, deviations could indicate that significantly different concentrations of each sample hybridised to the array were used.

SNP call rate

This metric measures the proportion of SNPs that have been successfully assigned a genotype.

SNP trough:peak ratio

High-quality SNP data for most samples results in three peaks in the log ratio distribution plot, while low-quality data may result in less well-defined peaks or even the complete loss of this expected distribution. As this may happen without affecting the SNP Call Rate, an additional metric to quantify the presence of this distribution has also been added — the SNP trough:peak ratio.

Spike-ins

This metric will either display "Match" or "No Match" depending on whether the Spike-ins associated with the sample (and only those Spike-ins) are identified within the data (i.e. they exceed the Spike-in Ratio Threshold — see **Tools -> Sample Identification**).

Standard deviation

The standard deviation of the log ratios of all mapped probes. Inferior to DLR spread as a quality control metric but included to enable comparison with legacy data.

Waviness

Some poor-quality samples present an undulating log ratio profile along the chromosome, making it more difficult to identify genuine aberrations in the data. However, the presence of these “waves” is not usually highlighted by standard quality control metrics, so a "waviness" metric can be calculated to indicate their presence and magnitude.

Quality Metric	Excellent Threshold	Satisfactory Threshold	Poor Threshold
X-Separation	N/A	N/A	N/A
DLR Spread	<0.2	<0.3	>=0.3
Green Background Noise	<15.0	<20.0	>=20.0
Green Signal Intensity	>1000.0	>800.0	<=800.0
Green Signal Intensity (SNPs)	N/A	N/A	N/A
Green Signal Reproducibility	<0.1	<0.2	>=0.2
Green Signal-to-Noise Ratio	>100.0	>30.0	<=30.0
Grid Alignment	N/A	N/A	N/A
Homozygosity	N/A	N/A	N/A
Negative Controls (Green)	N/A	N/A	N/A
Negative Controls (Red)	N/A	N/A	N/A
Non-Uniform Features	<0.5	<1.0	>=1.0
Red Background Noise	<25.0	<35.0	>=35.0
Red Signal Intensity	>1000.0	>800.0	<=800.0
Red Signal Intensity (SNPs)	N/A	N/A	N/A
Red Signal Reproducibility	<0.1	<0.2	>=0.2
Red Signal-to-Noise Ratio	>100.0	>30.0	<=30.0
Saturated Features	N/A	N/A	N/A
Signal Intensity Ratio	N/A	N/A	N/A

Quality Metric	Excellent Threshold	Satisfactory Threshold	Poor Threshold
SNP Call Rate	>0.988	>0.983	<=0.983
SNP Trough:Peak Ratio	<0.6	<0.8	>=0.8
Spike-ins	N/A	N/A	N/A
Standard Deviation	<1.0	<1.0	>=1.0
Waviness	<0.025	<0.0275	>=0.0275

Table 1: Default quality metric thresholds.

View sample quality metrics

There are several different ways to view the quality metrics for a sample, as listed below.

- **Genomic View Tab** — When viewing data in the **Genomic View** tab, quality metrics are displayed in the **Sample Information** box in left-hand panel. See Figure 27 for an example of a **Sample Information** box. If there are any potential issues with the quality, a yellow triangle containing an exclamation mark (⚠) will be displayed at the bottom of the box. One left-click on the triangle will bring up a message box (like that in Figure 28) containing information on which quality metrics pose issues.

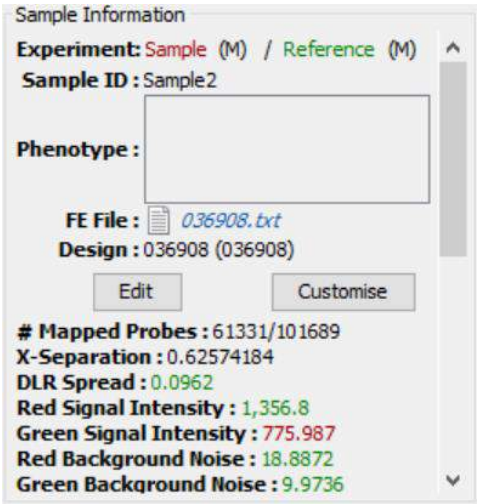


Figure 27: Sample Information box from the left panel of the Genomic View tab.

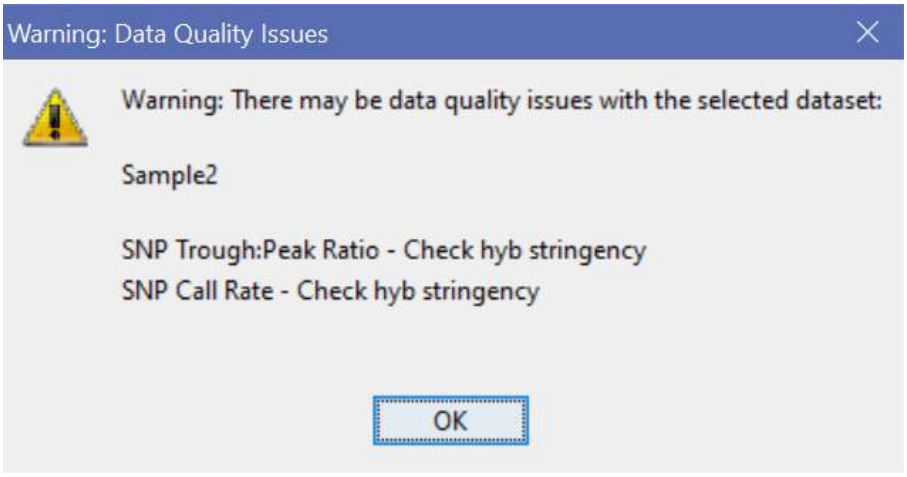


Figure 28: Data Quality Warning box detailing three potential issues with the data from Sample 2.

- **Database Management Tab** — It is also possible to view sample quality metrics from the **Database Management** tab. This option offers the additional functionality of plotting QC metrics across a number of samples. In order to do this, all sample data must first be submitted to the database. After doing this, navigate to the **Database Management** tab, as seen in Figure 29. It is then possible to select all samples you want to view using the tick boxes in the left table (which displays sample ID and gender by default). Then select **QC Metrics** from the central bar and a table should appear on the top right, where the first column represents QC metrics and every other column a selected sample. At the left side of this table, tick boxes are used to select which metrics to plot. Plots are displayed on the right-hand side underneath the metrics table (e.g. the DLR spread graph in Figure 29). Multiple graphs can be plotted simultaneously, but have to be viewed one at a time by clicking on the tabs at the top of the displayed graph.

  The easiest way to save data to the database is to navigate to the **Genomic View** tab and locate the sample specific vertical bar on the far left of the page. Click the lower save button ( ) to save the sample (or any updates to the sample) to the database.

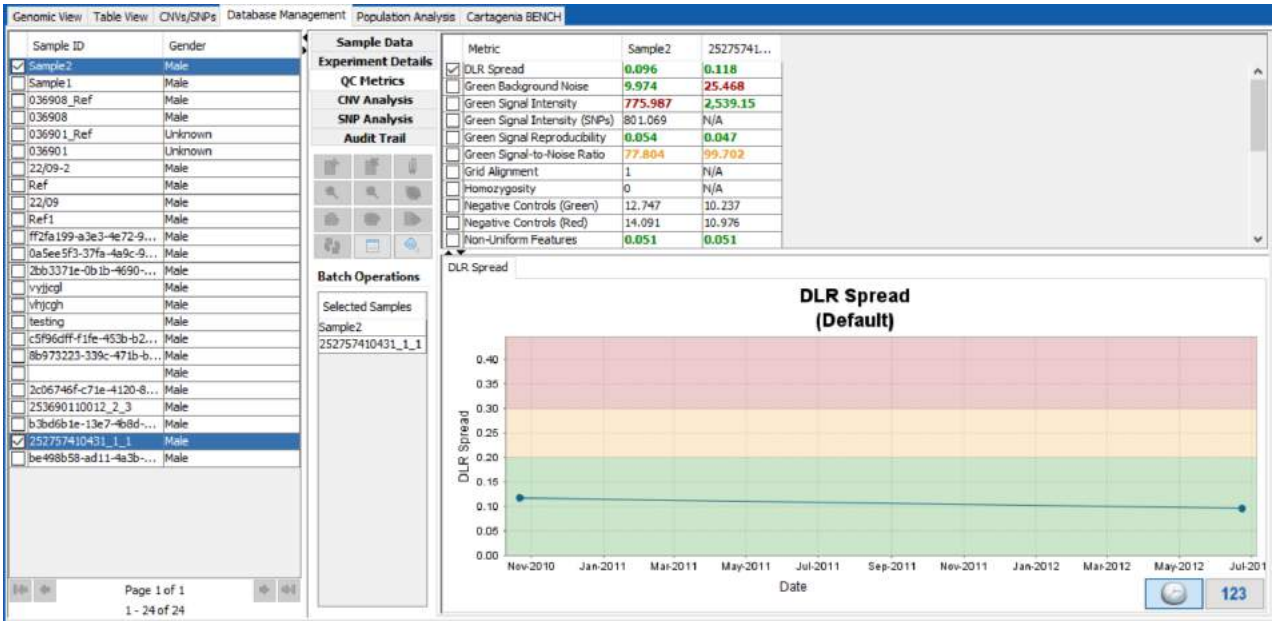


Figure 29: Database Management tab showing QC metric view.

- **Reports —** There are three types of report which the software can generate: PDF, HTML and tab-delimited text format. All of these reports list QC metrics near the beginning. See the Report generation section for help creating reports of your data.
- **FE file QC metrics —** It is possible to view some of the QC metrics (those contained in the FE files) before any analysis has taken place. In order to do this, choose **Tools** from the menu bar and then click **Extract QC Metrics from FE Files.** This will bring up a window like that seen in Figure 30. Choose **Add Files(s)** to load FE files into the table. The QC metrics contained within the FE files are as follows: Non-uniform features, DLR spread, green (and red) signal reproducibility, green (and red) background noise, green (and red) signal-to-noise ratio and green (and red) signal intensity.

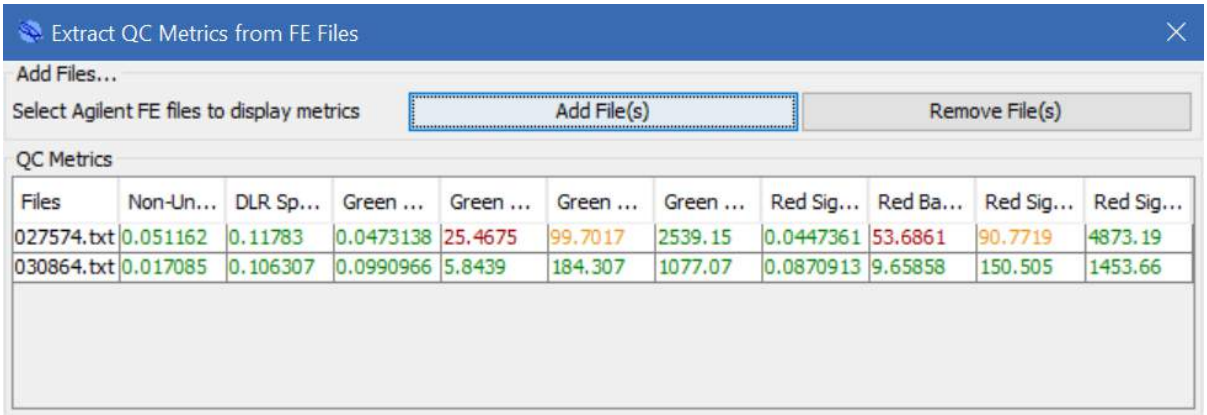


Figure 30: Extract QC Metrics from FE Files window.

Thresholds

QC metric thresholds are important to maintain consistency across analyses and to help the user identify problems with their data. The software usually has pre-defined and recommended default thresholds as well as different threshold sets for different arrays (e.g. 1x1m, 2x400k). Additionally, it is possible to alter the thresholds and to create new sets. For more details on each of these functionalities, please see the sections below:

Viewing thresholds

There are two main ways in which you can view QC metric thresholds and defined sets:

1. Choose **Tools -> Options... -> QC Metrics** to open the QC metric settings window, as seen in Figure 31. All defined threshold sets are listed along the left-hand side of the window, with the blue bar indicating which set is currently being displayed on the right-hand side (which is the *Default* set in Figure 31). To view metrics from a different set, click on the set in the list. When viewing non-*default* sets, all threshold values in bold are those which differ from the *Default* set.
2. These details can also be accessed from the **Database Management** tab. First, select **QC Metrics** from the central panel (see Figure 32) and then choose the settings icon () which can be found in the middle of the lower row of icons. This will bring up a **Metric Settings** window like that displayed in Figure 33. The functionality of this window mirrors that of the **QC Metrics** tab of the **Options** menu, as seen in Figure 1 and described in step 1.

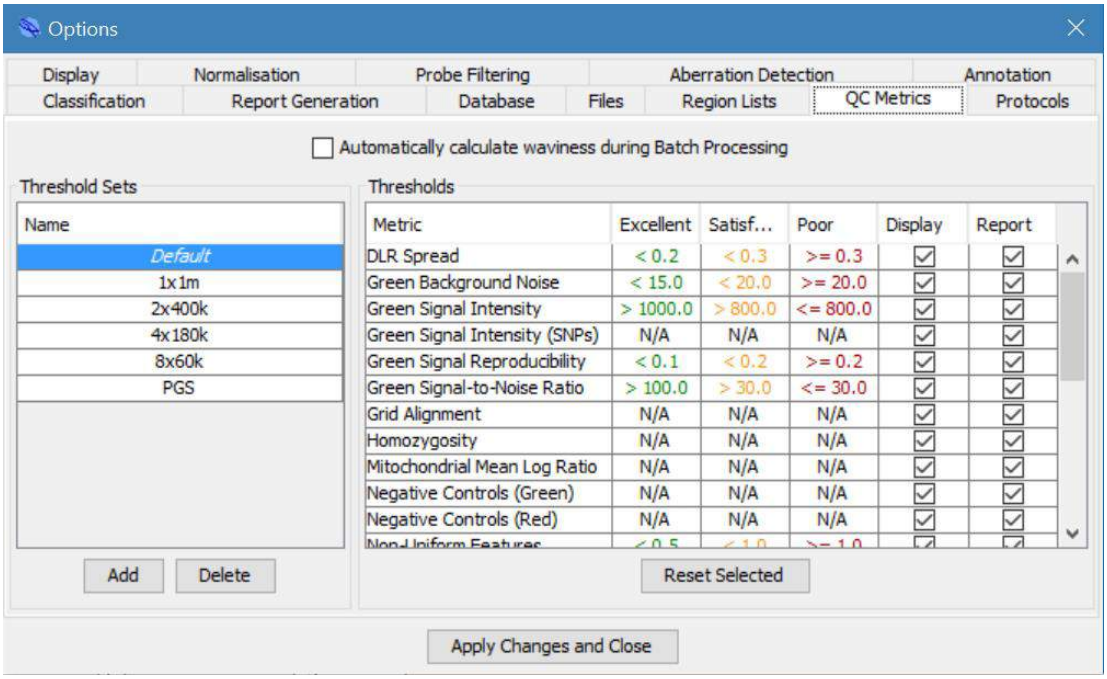


Figure 31: QC metrics setting panel.

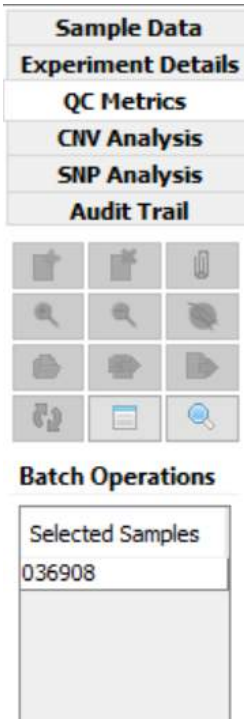


Figure 32: Central panel of Database Management tab indicting the location of the QC Metrics option.

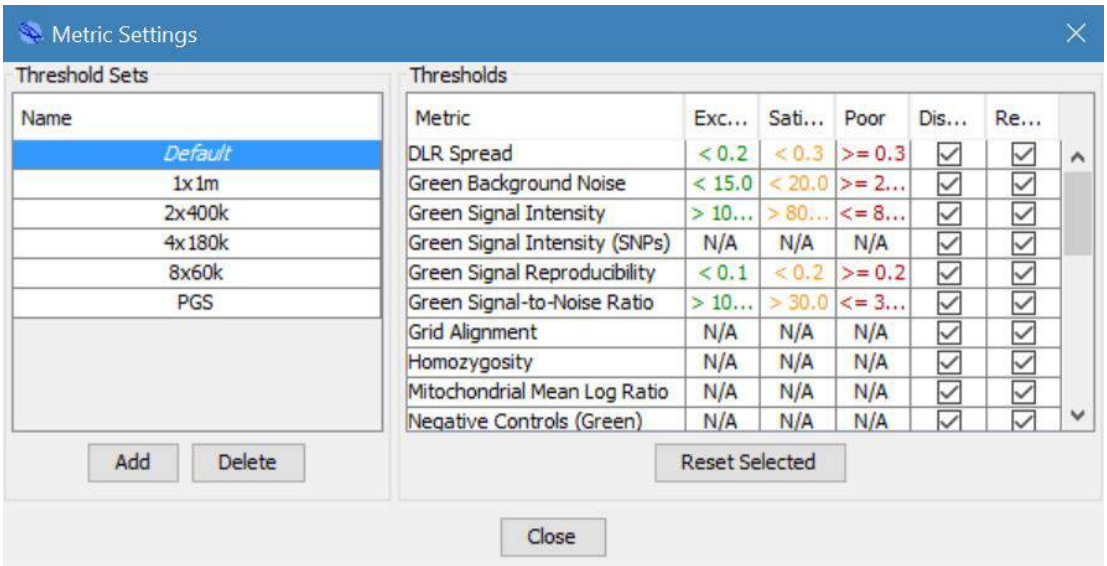


Figure 33: Metric Settings window which can be accessed through the Database Management tab.

Specifying threshold sets

Thresholds are applied to the data during the initial data processing stage. The threshold set is specified by the protocol loaded, or, if no protocol is being used, by the set chosen in the **Options** menu. To access this menu, choose **Tools -> Options... -> Protocols**. An example can be seen in Figure 34. If a protocol is loaded, the associated metric set is displayed in the **Metric Set** drop-down box. By choosing a new set from this drop-down box, the current protocol is removed, and it will use all current settings (which should be the same as the protocol if nothing is edited) including the new metric set. Alternatively, a different metric set can be chosen by loading a new protocol. This can be done by choosing an available protocol from the list and clicking **Load Protocol**, or by creating a new protocol using the **Create Protocol** button. Finally, choose **Apply Changes and Close** to close the window ready to begin analysis using the new settings.

Alternatively, when a sample file is imported for analysis the **Sample Details** window will be shown, from which it is possible to select a metric set for analysis, as seen in Figure 35. Selection of a metric set will override any loaded protocol setting, resulting in the protocol being removed, but maintaining all other protocol settings. The new set of settings will remain until changed manually.

For more information on importing feature extracted files please refer to the section Import feature extracted files.

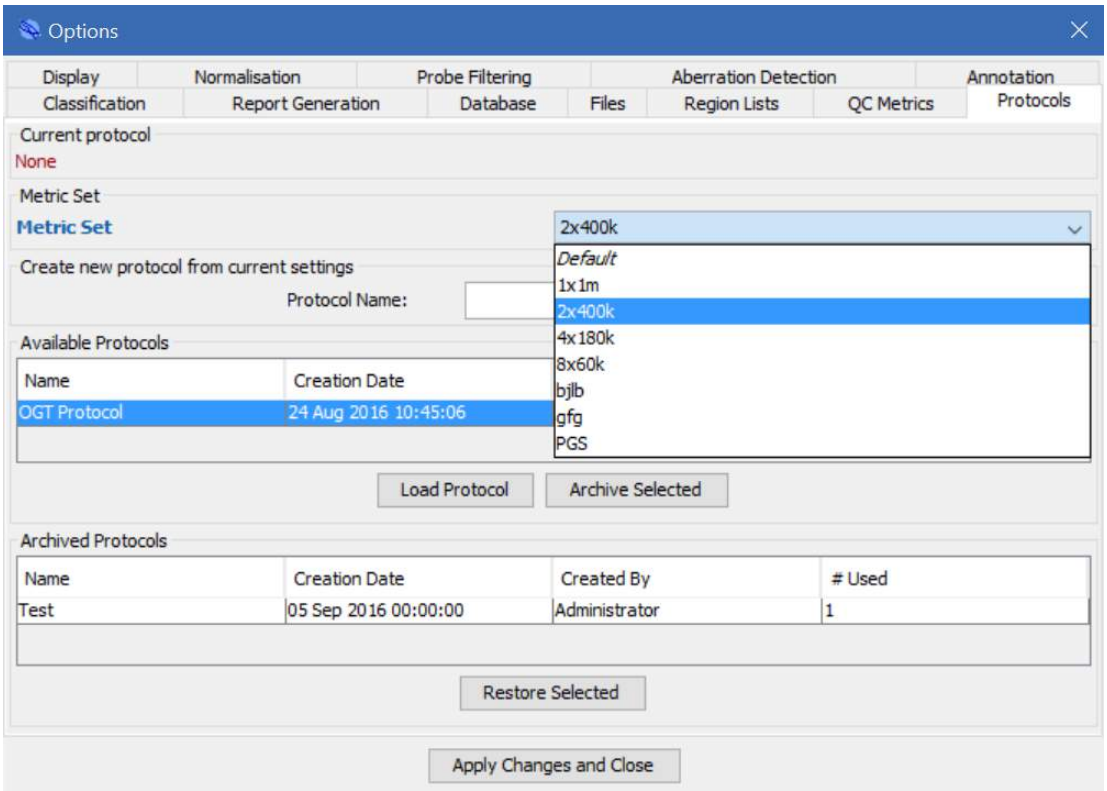


Figure 34: Protocols tab of Options menu.

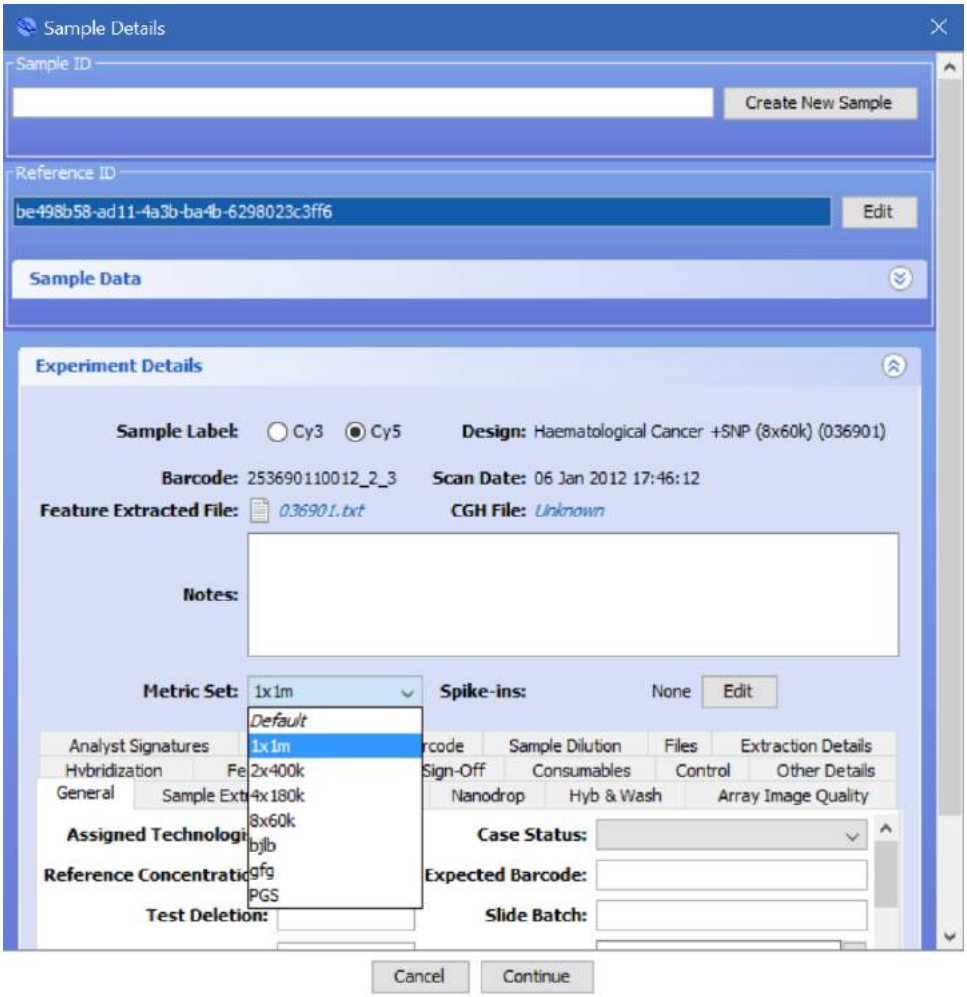


Figure 35: Specifying a metric set from the *Sample Details* window.

Editing and creating thresholds

Editing thresholds and creating new threshold sets can be carried out from either of the settings menus (Figures 31 and 33). In order to open each of these menus, please follow the steps outlined in the View thresholds section. To edit a threshold for an existing set, double left-click on the value of interest to make the box editable. Type the new value and then single left-click elsewhere in the window to load the new threshold. Click **Apply Changes and Close** or **Close,** depending on which window you are in. Any displayed QC metrics should re-load to use the new thresholds.

The software will not give a warning message if you enter a non-numerical value as a threshold. If, when you can left-click outside the entry box, the cursor is removed and the entered value displayed in colour, then the new threshold has been accepted by the program. If the cursor remains in the entry box, then the value entered is invalid.

In both of the options windows there is a button labelled **Reset Selected**. This option can be used on bold values (those which differ from default) to revert them to default values. To do this, left-click on the row of thresholds in the table which you would like to revert (and where the values are bold), then click the button. A window will appear asking if you are sure you want to revert this QC metric to default settings.

Furthermore, it is possible to create new threshold sets by choosing **Add** (see Figures 34 and 36). This will load a box into which the name of the new set is entered. The new set will appear in the list and contain all default values. To edit the threshold values, follow the guidance at the beginning of this section.

Edit aberrations

Whilst CNVs and LOH regions can be automatically detected by CytoSure Interpret Software, there are occasions in which manual editing of the aberrations is required. CytoSure Interpret Software boasts functionality to add, remove and merge aberrations, as well as adjust region boundaries and edit aberration details.

Adding aberrations

To manually add a CNV, it is necessary to select the whole region of interest (and nothing else) in the **Chromosome Section** of the **Genome View** tab. Then navigate to the **CNV Regions** table (in the lower left of the screen) and press the **Add** button. This will create a new aberration in the displayed region, with the aberration boundaries at the outermost displayed probes, as seen in Figure 34. The aberration will be identified as a gain or a loss depending on the mean log ratio of the included probe.

The same procedure can be used to add an LOH region by instead selecting the region in the **Allele Status** or **B-Allele Frequency** plot and pressing the **Add** button in the **LOH Regions** table.

To select a region in the **Allele Status** and **B-Allele Frequency** plots, SNP data zooming must be enabled. To do this, right-click on the plot and select **Enable SNP Data Zooming**.

Figure 34: Manually added aberration showing region boundaries at the outermost probes displayed.

Removing aberrations

There are two ways in which to delete an aberration:

1. Navigate to the **CNV** (or **LOH) Regions** table found at the lower left corner of the **Genomic View** page (as seen in Figure 34). Select the aberration of interest in the table and click **Remove.**
2. Locate the CNV in the **Chromosome Section** graph. Right-click and choose **Delete.** This is only possible for CNVs and not LOH regions.

Any change to CNV or LOH detection settings that triggers a recalculation of aberration may result in a deleted region re-appearing.

Merging CNVs

To merge CNVs, edit the **Chromosome Section** (of the **Genome View** tab) so that all CNVs to be merged (and no other CNVs) are displayed. Then choose **Tools** from the menu bar and in the drop-down menu select **Merge Displayed Aberrations.**

Region Adjustment

Further flexibility in aberration detection is achieved through manual adjustment of CNV and LOH region boundaries via the **Chromosome Section** and **Allele Status** / **B-Allele Frequency** plots in the **Genomic View**. To make such adjustments, hover the mouse over the edge of an aberration so that the default cursor is replaced with the resize cursor and drag the region to the required start/end location. An example of dragging the boundaries of an aberration is displayed in Figure 35.

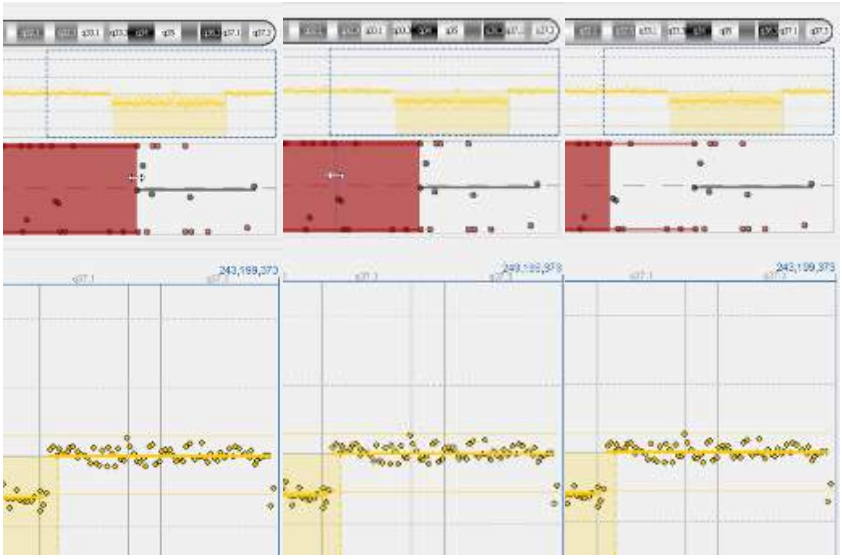


Figure 35: Manual adjustment of LOH region boundaries.

Editing aberration details

The **Edit Aberration Details** window, as seen in Figure 36, can be used to specify or edit the following details (for CNV aberrations only):

- Copy number
- Classification (See Manual classification for more details)
- Inheritance (bi-parental, maternal, paternal or *de novo*)
- Confirmation method
- Comments
- % Mosaicism

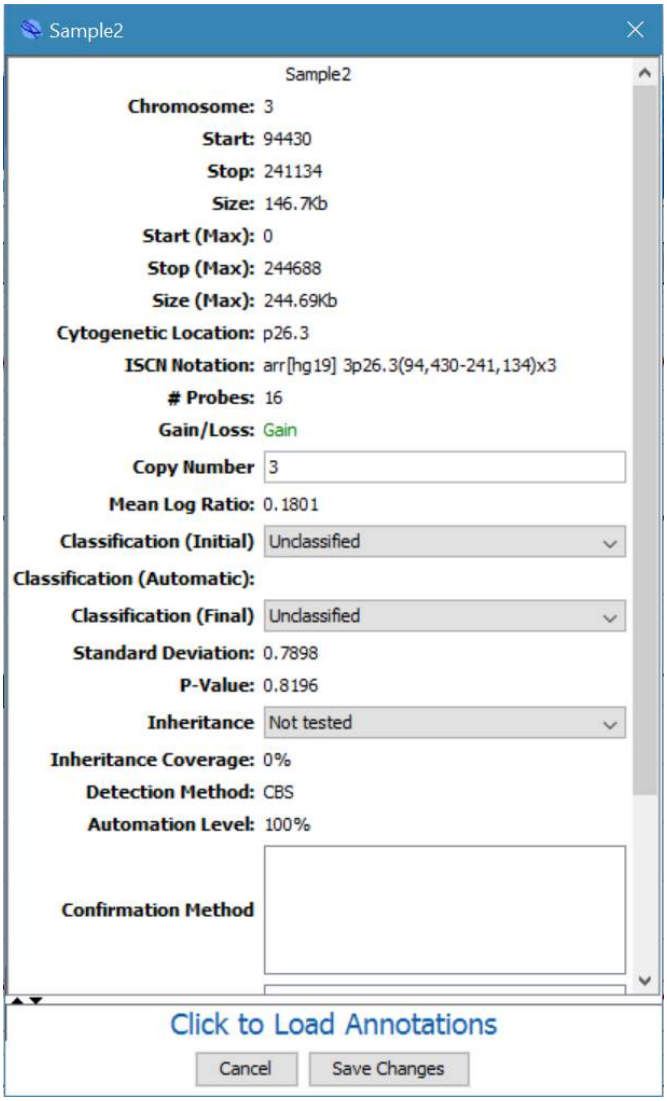


Figure 36: Edit Aberration Details window.

This window can be accessed in several ways:

1. Through the **Genomic View** tab
    a. Right-click on the aberration in the **CNV Regions** table and choose **Edit Aberration Details** (Figure 37).
    b. Double left-click on the aberration in the **CNV Regions** table.
    c. Locate the aberration in the **Chromosome Section** plot. Right-click on it and choose **Edit** from the drop-down menu (Figure 38).
2. Through the **CNVs/SNPs** tab

a. The left-hand section of the page, labelled **Aberration Details,** is the same as the **Edit Aberration Details** window**.** Select an aberration in the **Detected CNVs/SNPs** section to load its details into the **Aberration Details** section. Aberrations can be selected by clicking on them in the table or the ideogram.

b. When in the **Table View**, double left-click on an aberration in the table to bring up the **Edit Aberration Details** window.

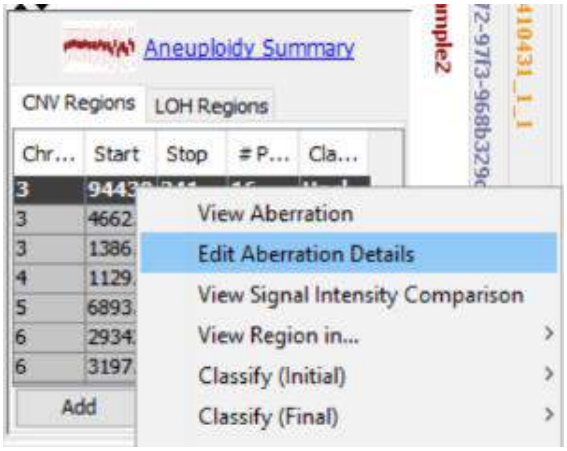


Figure 37: Edit Aberration Details window access via the CNV Regions table.

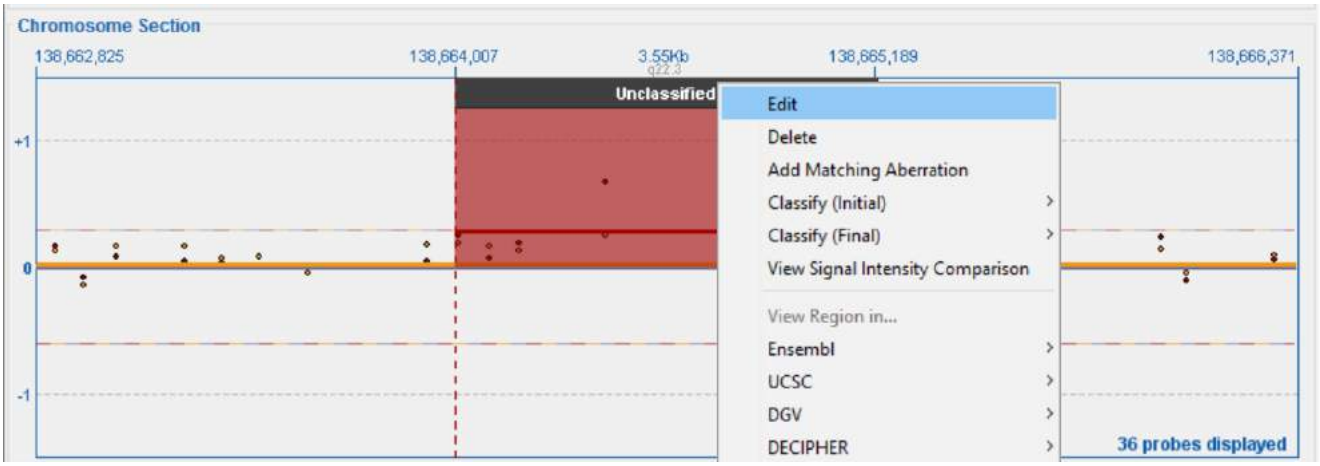


Figure 38: Edit Aberrations Details window access via the Chromosome Section plot.

Aberration classification

Following the identification of CNVs and regions of LOH, assessment of their phenotypic impact can be carried out through aberration classification. This will help to filter out unimportant aberrations and highlight those which may be of greater scientific interest. The ability to specify an initial and a final classification can provide the means for different

researchers to collaborate on the classification process, as well as highlight level of certainty and encourage the process of additional curation.

Classification categories

A set of standard classifications are provided with the software, as listed here:

- Benign
- Uncertain significance: likely benign
- Uncertain significance
- Uncertain significance: likely pathogenic
- Pathogenic
- Ignore

Additionally, customised classifications can be generated through the **Classification** tab of the **Options** menu (Figure 39). To access this, select **Tools -> Options... -> Classification**. The **Existing classifications** section displays all currently defined classifications; all those which are available to use in data analysis. The **Add new classification** section provides the functionality to add, remove, archive or restore classification categories. Addition of new categories allows a user or lab to specialise the analysis. In some cases, classifications used by a lab to label detected CNVs have changed as analysis and interpretation processes have developed, resulting in an unwieldy list of classification types, many of which are obsolete. In these circumstances, it may be useful to remove classifications from the system altogether. However, removing classifications requires re-assigning affected CNVs with one of the remaining classifications, which may not be appropriate. To enable maintenance of a manageable list of classification types without the loss of historical information, classifications can also be archived. An archived classification will be retained by the system but will not be available for assignment to CNVs during the interpretation process. To move an archived classification back to the **Existing classifications** table, the restore function is also available. Steps required to carry out these processes are outlined below:

Add a classification:

1. Type the name of the new classification category into the **Name** entry box (in the **Add new classification** section of **Classification** tab)

2. Choose an associated colour by clicking on the colour swatch (black square in Figure 39) and then selecting the colour of choice from the menu (for more details on colours see Classification colours).
3. Click the **Add Classification** button to create the classification. The newly generated category should subsequently appear in the **Existing classifications** table below.

Remove a classification:

1. Select the classification of interest from either the **Existing classifications** or the **Archived classifications** tables by left-clicking on the row, which will highlight it blue. Multiple classifications can be selected by clicking and dragging.
2. Click the **Remove Selected** button to delete the classification(s).

Archive a classification:

1. Select the classification of interest from the **Existing classifications** table by left-clicking on the row, which will highlight it blue. Multiple classifications can be selected by clicking and dragging.
2. Click the **Archive Selected** button to move the classification(s) from the **Existing classifications** to the **Archived classifications** table.

Restore a classification:

1. Select the classification of interest from the **Archived classifications** table by left-clicking on the row, which will highlight it blue. Multiple classifications can be selected by clicking and dragging.
2. Click the **Restore Selected** button to move the classification(s) from the **Archived classifications** to the **Existing classifications** table.

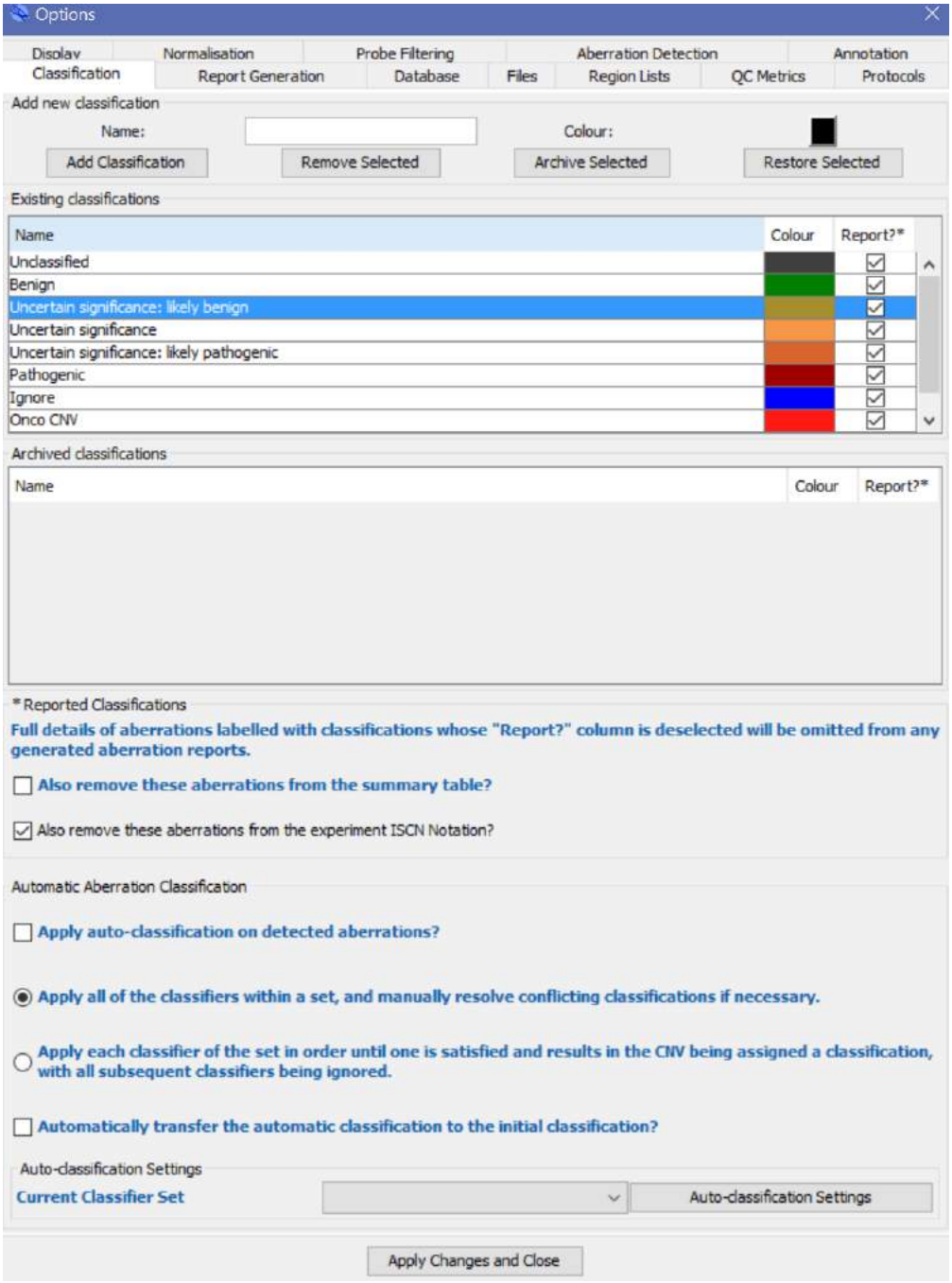


Figure 39: Classification tab of Options menu from which classifications can be added, removed and edited.

Classification colours

Each classification category is assigned a colour, which is used throughout the system to aid visual analysis. These colours can be found in the following places:

- On the **Genomic View** tab:
    - The **Chromosome Section** plot will display a bar at the top of every CNV which has the text description of the assigned classification and is shaded in the associated colour for that classification. An example of this can be seen in Figure 39 where one CNV is classified as **Uncertain significance: likely benign**, which is associated with a mustard colour, and the other as **Benign**, which is associated with a green colour. Please note that the red colour shown in Figure 39 represents the colour of the sample and not the CNV classification.
    - The **B Allele Frequency / Allele Status** plot will shade the whole section of a detected LOH region with the classification colour (or the sample colour if no classification has been given). Figure 40 shows an LOH which has been classified as **Benign** and is therefore displayed in green (the colour associated with the **Benign** classification).
    - The **CNV/LOH Regions** tables shade the colour of each row with the associated classification. If the aberration is unclassified then the row will be grey. See Figure 39 for an example.

The final classification colour will always have precedence over the initial classification colour. Therefore, any observed aberration colour will be that of the final classification, unless only an initial classification is given, in which case it will be the colour associated with that.

Figure 39: Lower section of Genome View tab displaying CNV classification colours

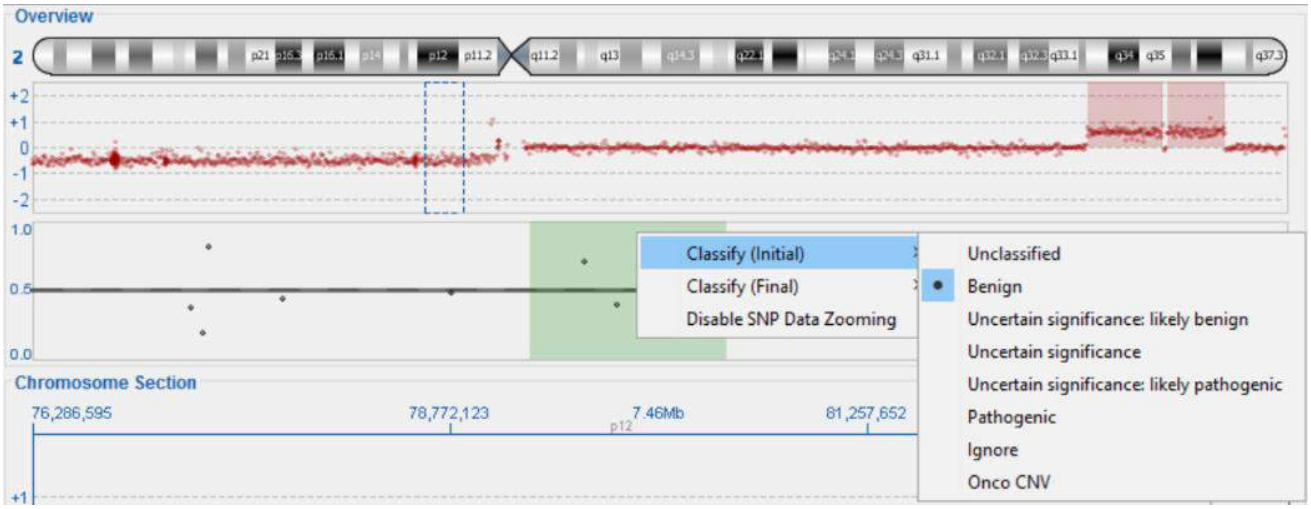


Figure 40: Upper section of Genome View tab showing LOH classification colour on the B Allele Frequency plot.

- On the **CNVs/SNPs** tab:
  - The **Detected CNVs/SNPs** table will show colours in the same way as the **CNV/LOH Regions** table on the **Genomics View** tab. That is, the colour of each row will represent the associated aberration's classification; the initial classification if no final is specified, otherwise the final classification. Unclassified aberrations will be coloured in grey. See Figure 41 for an example.

The colours of the CNV bars alongside the chromosomes in the **Ideogram View** of the **CNVs/SNPs** tab represent the gain (green/blue) or loss (red) status of the aberration and not the classification.

Detected CNVs/SNPs

Chromosome	Start	Stop	# Probes	Classification (Initial)
2	204598942	219263558	92	Uncertain significance: likely benign
2	219992599	231235872	71	Benign
7	61081314	62353488	9	Unclassified
7	98789198	98856948	4	Unclassified
12	37896066	38276243	5	Unclassified
13	28608193	28608262	18	Unclassified
13	28608340	28608436	21	Unclassified
14	19457944	20420346	7	Unclassified
14	106057624	107286764	48	Unclassified
15	20003669	28704700	47	Unclassified
15	34533380	42636858	51	Unclassified
15	34533380	43438540	56	Unclassified

Figure 41: Detected CNVs/SNPs table of CNVs/SNPs tab showing classification colours.

- On the **Database** tab:
    - As shown in Figure 42, the **CNV/SNP Analysis** table also displays colours in the same way as the **CNV/LOH Regions** table on the **Genomics View** tab. That is, the colour of each row will represent the associated aberration's classification; the initial classification if no final is specified, otherwise the final classification. Unclassified aberrations will be coloured in grey.

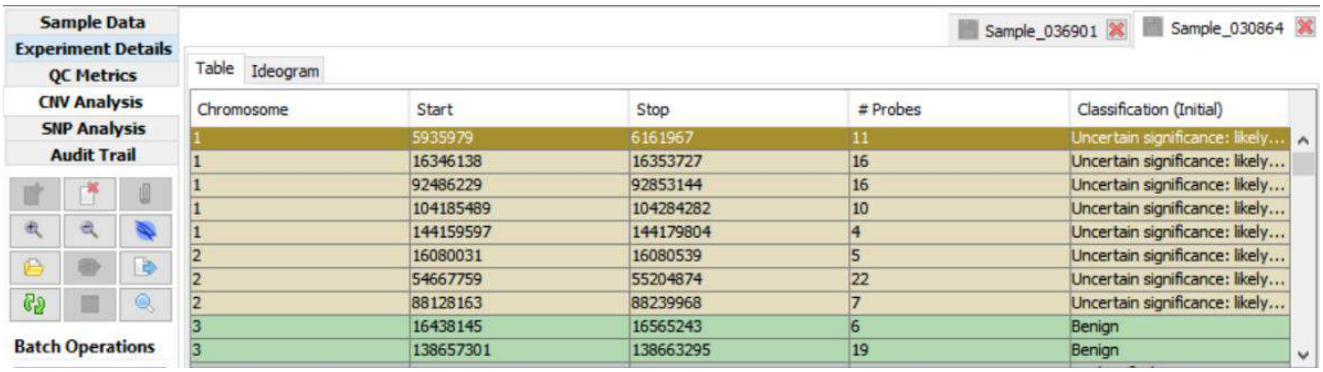


Figure 42: CNV Analysis table of the Database Management tab displaying classification colours.

Manual classification

You can classify aberrations in various ways:

1. From the **Genome View** tab
    a. Right-click on an aberration (or a group of aberrations) in the **CNV/LOH Regions** table (lower left corner) and select **Classify (Initial/Final)** from the option menu, followed by the classification of choice.
    b. Right-click on an aberration in the **Chromosome Section** or **Allele Status / B-Allele Frequency** plot and select **Classify (Initial/Final)** from the option menu, followed by the classification of choice. See Figure 43 for an example.
    c. Bring up the **Edit Aberration Details** window for a particular aberration and locate the **Classification** options. See the section Edit aberration details for ways to access this window.

2. From the **CNV/SNPs** tab
    a. Right-click on an aberration (or a group of aberrations) in the **Detected CNVs/SNPs** table (left hand side) and select **Classify (Initial/Final)** from the option menu, followed by the classification of choice. An example of this process is shown in Figure 44.
    b. Bring up the **Edit Aberration Details** window for a particular aberration and locate the **Classification** options. See the section Edit aberrations for ways to access this window.

3. From the **Database Management** tab
    a. If an aberration has been submitted to the database, right-click on the corresponding row(s) of the regions table (labelled **Table** or **LOH regions**) in the **CNV** or **SNP Analysis** section. This will allow selection of the options **Classify (Initial/Final)** and the subsequent classification of choice. An example of this can be seen in Figure 45.

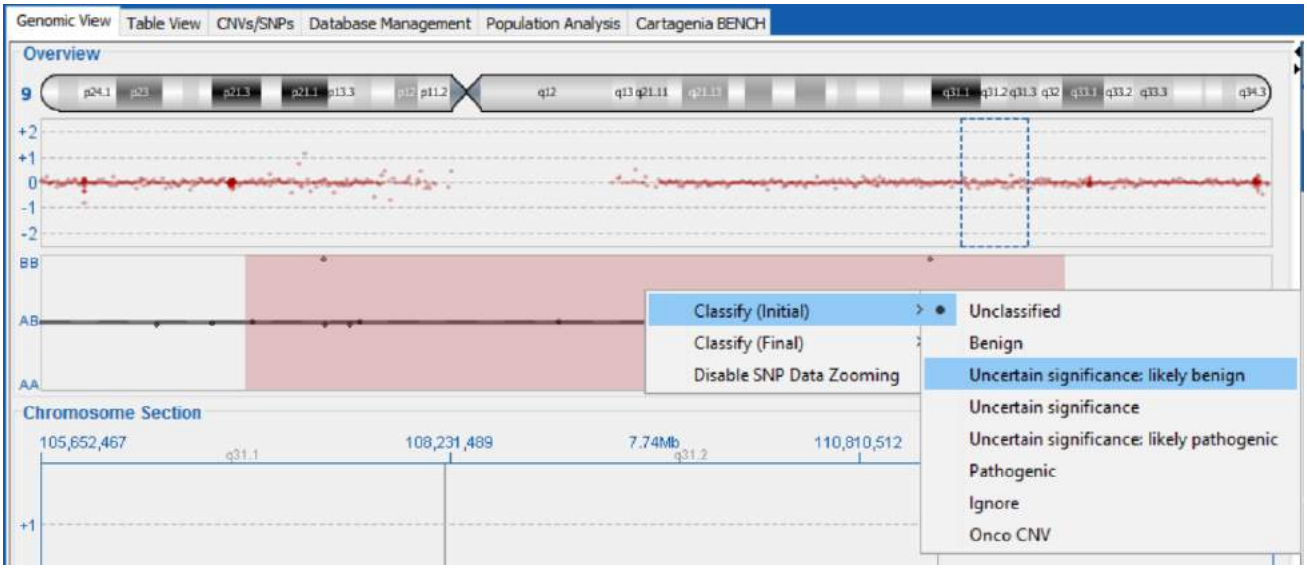


Figure 43: Classification of an aberration from the Allele Status plot.

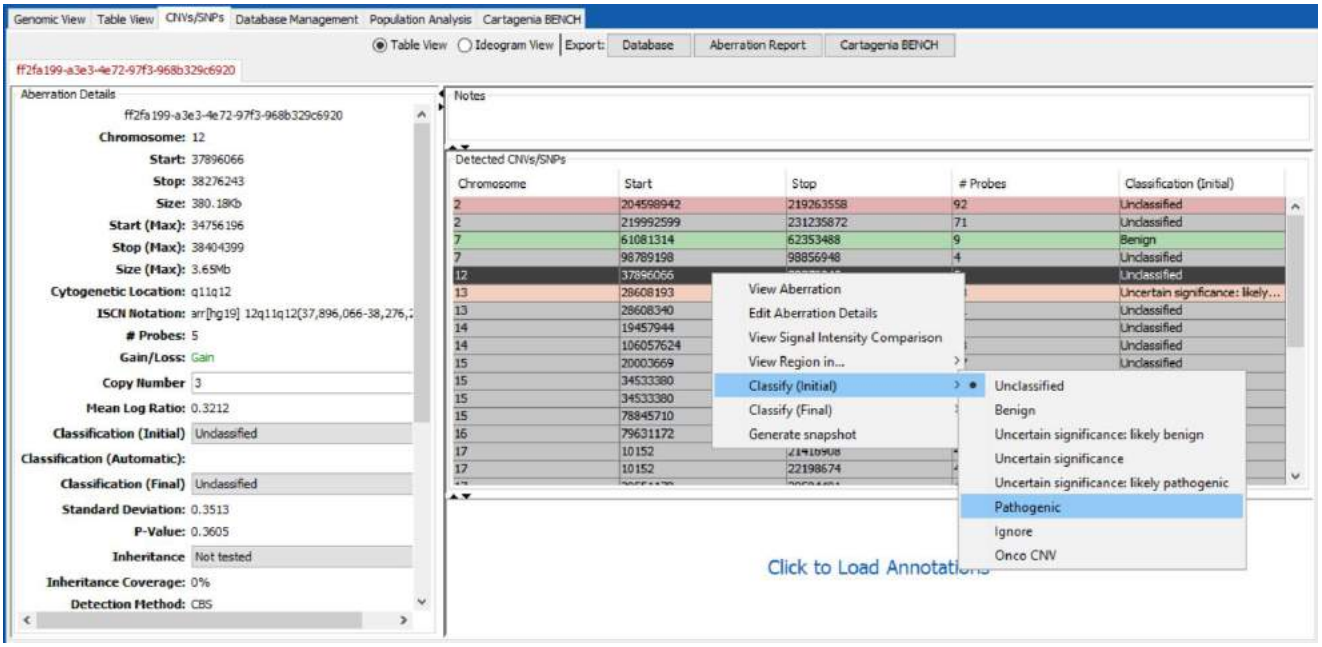


Figure 44: Classification of an aberration from the Detected CNVs/SNPs table.

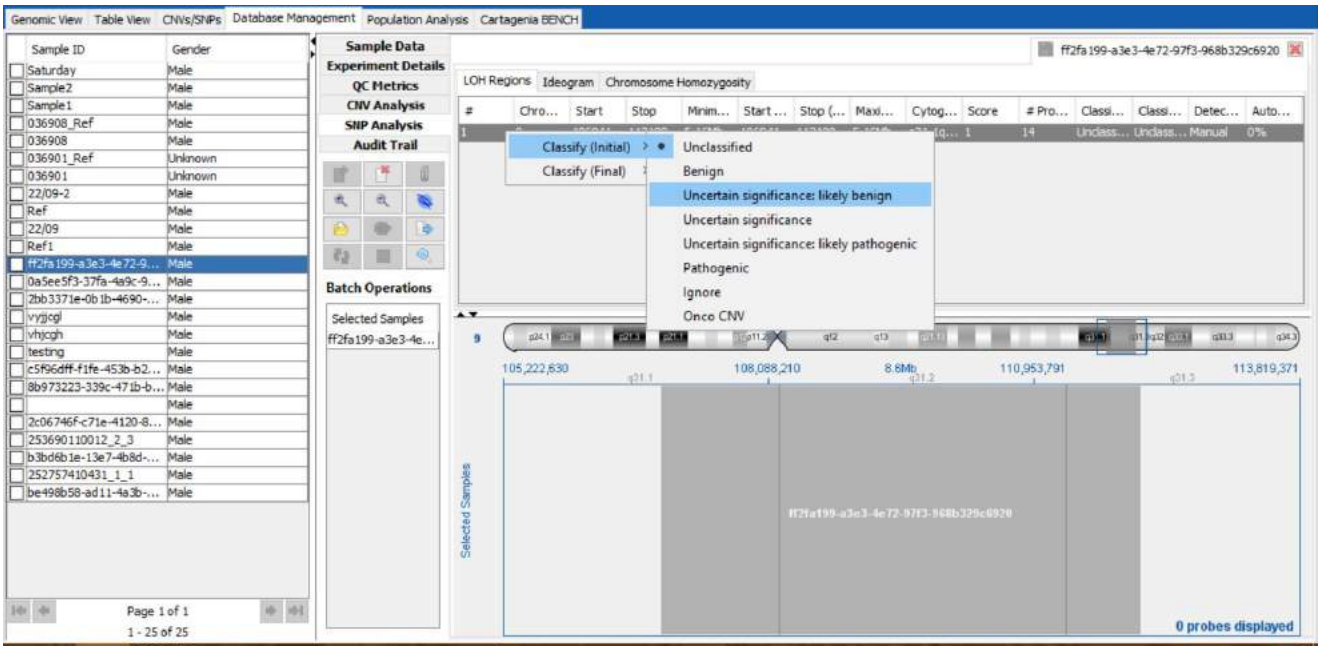


Figure 45: Classification of an aberration from the LOH Regions table of the SNP Analysis section of the Database Management tab.

Automatic classification

To enable automatic classification, users create a “Classifier Set” made up of multiple “Classifiers”. Each classifier describes the criterion against which a CNV will be tested and the classification that it will be assigned if it satisfies the criterion. For example, a user may create a classifier that checks whether a CNV is within an annotation in the Copy Number

Variation track and, if so, assigns the “Benign” classification to that CNV automatically. Once the set of classifiers is constructed, the user may opt to apply the classifiers in order until one passes its criterion, or apply all classifiers and resolve conflicting classifications manually if necessary. Multiple classifier sets can be created for use in different types of analyses, and the selected classifier set can be linked to an analysis protocol in order to standardise interpretation when the protocol is in use.

The following sections describe the administration and application of the automatic classification functionality in full.

Classifier Set Creation

To enable automatic CNV classification, the user must first create a classifier set via **Tools -> Options... -> Classification -> Automatic Aberration Classification -> Auto-classification Settings -> Auto-classification Settings.** This will open the **Automatic Classification Settings** window containing the controls to manage the classifier sets and individual classifiers that are available to the user.

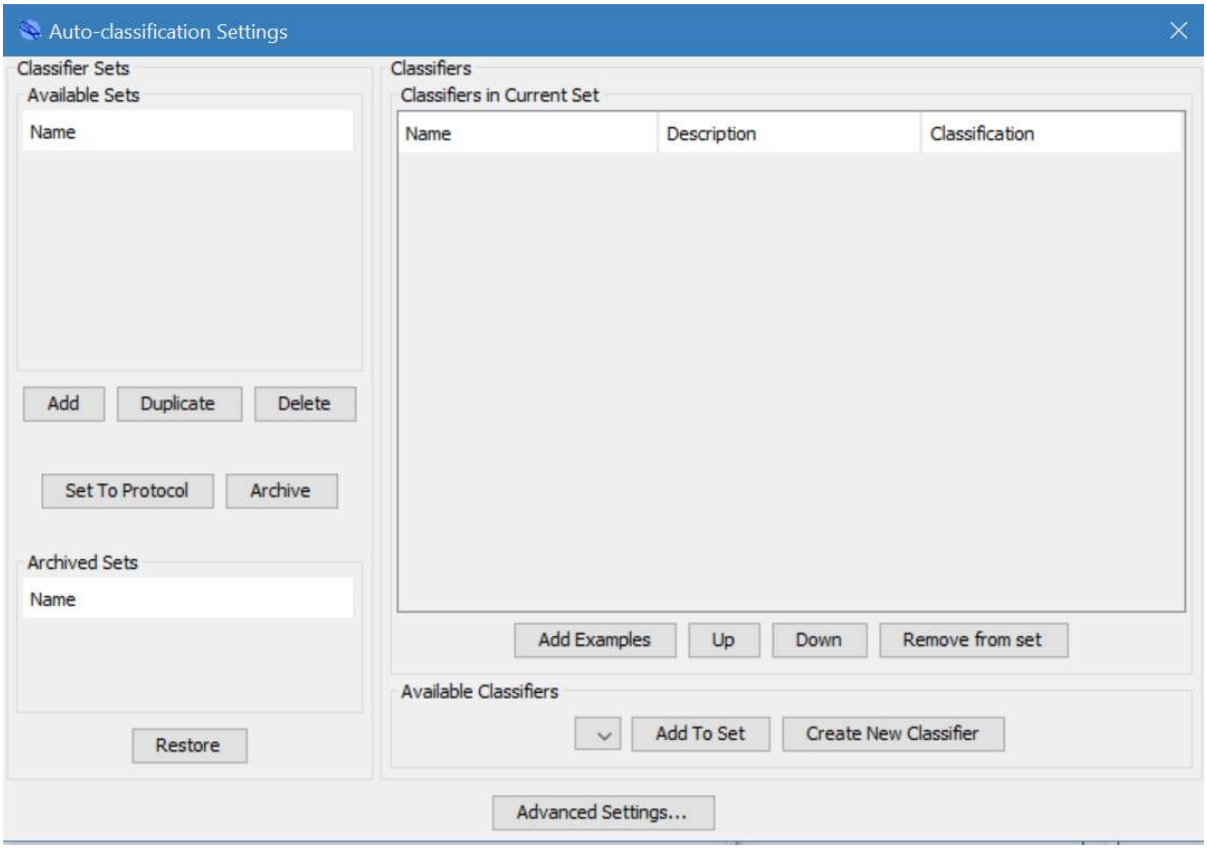


Figure 46: Auto-Classification Settings window providing functionality for configuration of classifier sets.

To create a new classifier set, click the **Add** button in the **Classifier Sets** section, enter a name for the new set and click **OK**. The name of the set will then appear in the **Available**

Sets table. To apply classifiers to a set, it is first required that classifiers are created, as described in the following section.

Classifier creation

To create a classifier, click the **Create New Classifier** button in the **Available Classifiers** panel at the bottom of the screen (Figure 46). This will open the **Create Classifier** window (Figure 47).

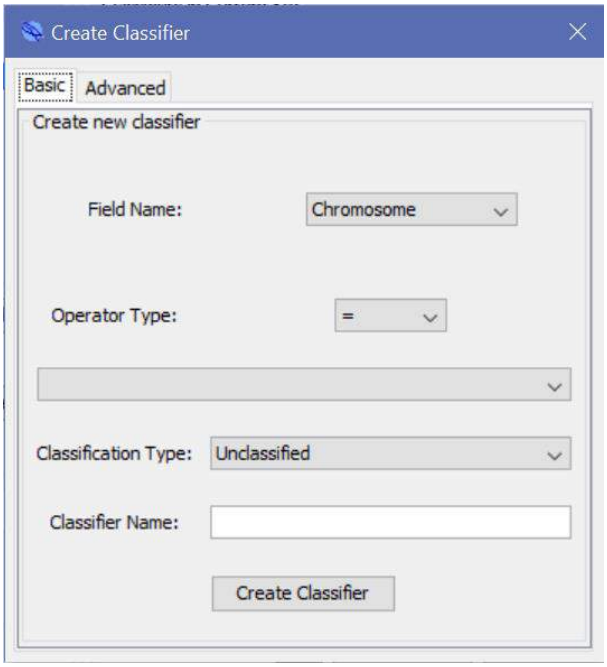


Figure 47: Create Classifier window providing the means to create new classifiers (Basic tab).

The software enables the user to create two types of classifier — **Basic** and **Advanced**.

Basic classifiers test a standard attribute of each CNV against the given criteria. For example, they can determine whether the CNV is a duplication, or whether the mean log ratio of the CNV exceeds a certain threshold, and assign a classification accordingly. To create a Basic Classifier, in the **Basic** tab (Figure 47) select the required **Field Name** and **Operator Type**, enter the value against which the CNV will be compared, select the **Classification Type** that will be assigned to a CNV that satisfies the criterion, enter a **Classifier Name** describing the Classifier and click the **Create Classifier** button.

Advanced Classifiers consider the annotations in a particular track within each CNV region, allowing the user to classify based on the names of these annotations, the number of annotations in the region, or the total proportion of the CNV region covered by the annotations (the “Overlap Statistic”). For example, users can create an Advanced Classifier to test whether a CNV covers an OMIM gene, or whether more than 80% of the CNV region is covered by annotations in the Copy Number Variation track, and assign a classification accordingly. To create an Advanced Classifier, in the **Advanced** tab (Figure 48) select the **Annotation Type** and the **Overlap Statistic** to calculate for the CNV region in the selected track, then select the **Operator Type** and enter the appropriate value into the field(s)

beneath to create the criterion. To complete the Classifier, select the **Classification Type** to which CNVs that satisfy the criterion should be assigned, enter a name for the Classifier in the **Classifier Name** field and click the **Create Classifier** button.

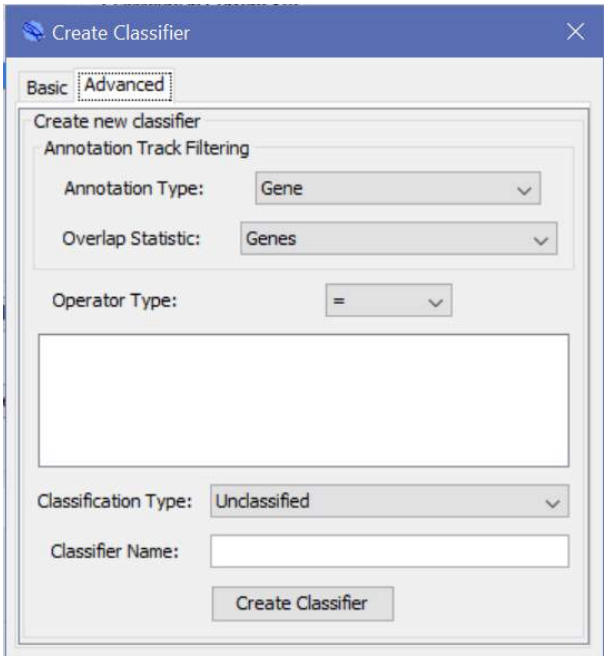


Figure 48: Create Classifier window providing the means to create new classifiers (Advanced tab).

Adding classifiers to classifier sets

All classifiers that have been created by the user (that haven't been archived) are listed in the **Available Classifiers** drop-down box in the **Auto-classification Settings** window (Figure 46). To add a classifier to a set, simply select the set from the list of **Available Sets**, select the classifier from the **Available Classifiers** drop-down box and click the **Add To Set** button. The selected classifier, which will appear in the **Classifiers in Current Set** table, can be individually moved up or down the list to determine the order in which they will be applied by using the **Up** and **Down** buttons and can be removed from the set by clicking the **Remove from set** button.

The **Auto-classification Settings** window also provides the means to add a list of example classifiers as a potential starting point for building a classifier set via the **Add Examples** button.

If any annotation tracks required by the classifiers have not been installed, or are not named as expected, the classifier may fail to be added.

Applying the classifier set

When the classifier set has been created, close the **Auto-classification Settings** window to return to the **Classification** tab in the **Options** window (Figure 49). The new set will appear in the **Current Classifier Set** drop-down box, and selecting it will ensure that it is used whenever the software runs automatic aberration classification. CytoSure Interpret Software can be configured to run automatic aberration classification whenever a CNV is detected by selecting the **Apply auto-classification on detected aberrations?** checkbox, and the user can also manually trigger application of the classifier set by selecting **Tools -> Apply Auto-Classification**.

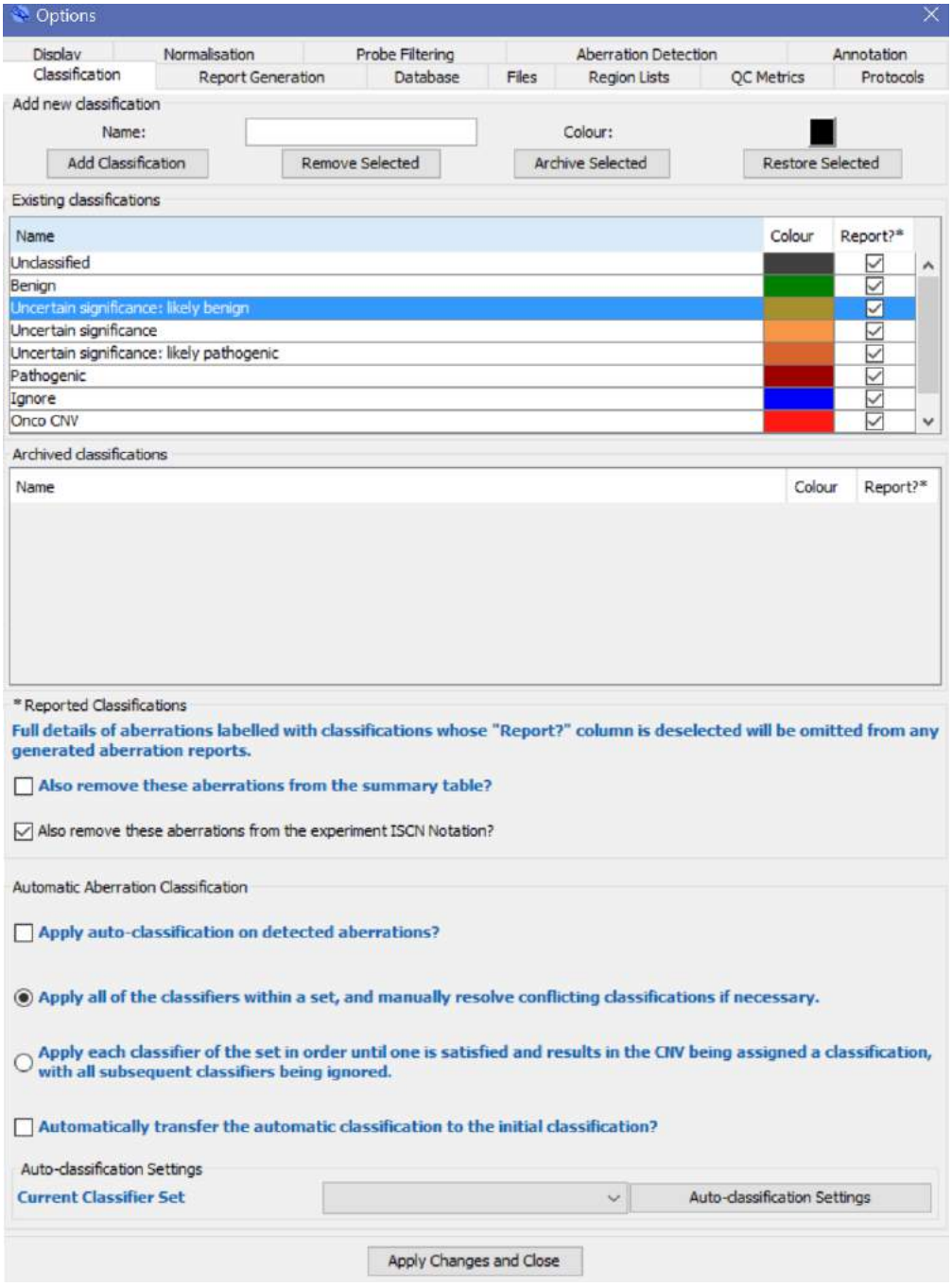


Figure 49: Classification tab of Options menu.

Please note that calculation of advanced classifiers may be accelerated by loading the relevant annotation tracks into memory by right-clicking on the track in the **Genomic View** and selecting **Load into Memory**.

The **Current Classifier Set** and the other settings in the **Automatic Aberration Classification** section described below are analysis protocol options, meaning that they are linked to the analysis protocol and their modification will result in the nullifying of the current protocol. This also means that it is possible to establish which classifier set (if any) was used in all experiments run in **Workflow mode.**

Application options

The user can choose to apply the classifier set in one of two ways:

1. Apply all of the classifiers within a set, and manually resolve conflicting classifications if necessary.
2. Apply each classifier of the set in order until one is satisfied and results in the CNV being assigned a classification, with all subsequent classifiers being ignored.

The user should select the appropriate option in the **Automatic Aberration Classification** section (Figure 50) to suit their requirements in this respect.

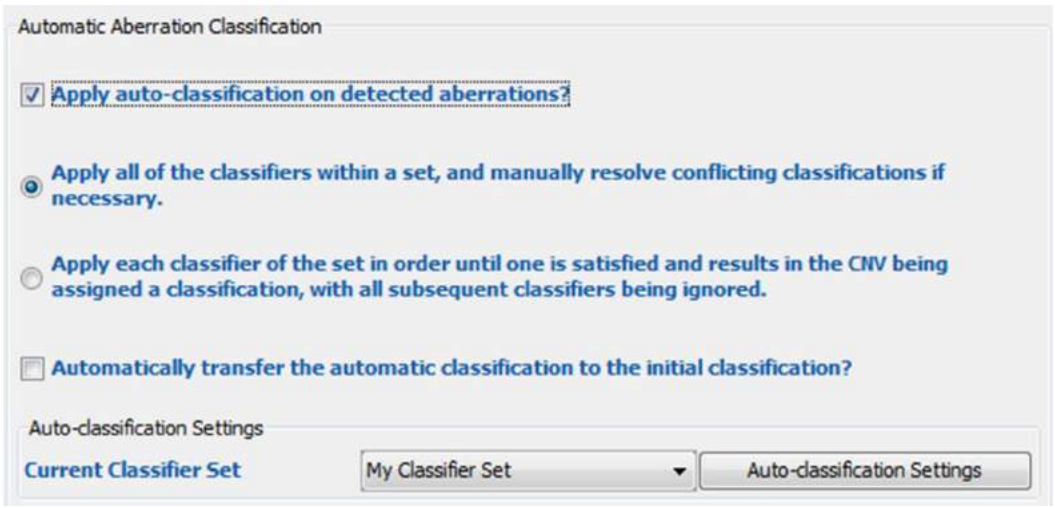


Figure 50: Automatic Aberration Classification section of the Classification tab of the Options window.

When a classification is assigned to a CNV via automatic aberration classification, the **Classification (Automatic)** field is set by the software to suggest an initial classification to the user. If required, users can configure the software to transfer the automatic classification to the **Classification (Initial)** field as part of the automatic aberration classification process by ensuring that the **Automatically transfer the automatic classification to the initial classification?** checkbox is selected.

Advanced settings

Further customisation of Classifiers is available via the **Advanced Settings...** button in the **Auto-classification Settings** window, where existing classifiers may be archived or translated into new classifiers to create more complex criteria.

In order to maintain a manageable list of classifiers, it is possible to archive classifiers that are no longer required so that they do not appear in the **Available Classifiers** drop-down box in the **Advanced Auto-classification Settings** window by selecting them in the **Available Classifiers** table and clicking on the **Archive** button (Figure 51). Archived classifiers will appear in the **Archived Classifiers** table and can be restored for subsequent use in a classifier set by clicking on the **Restore** button.

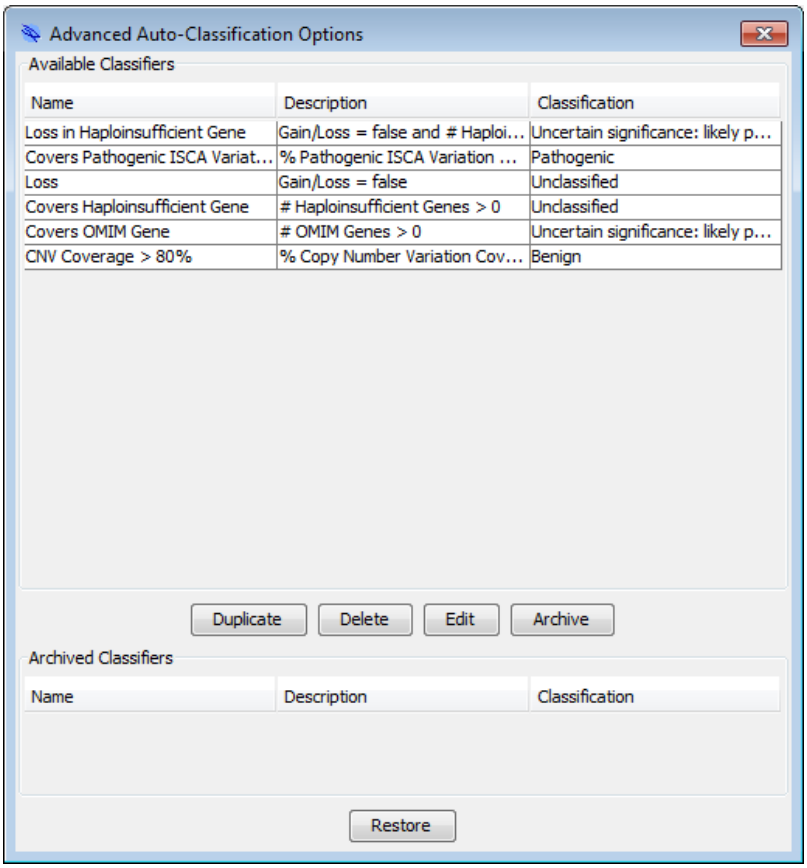


Figure 51: The Advanced Auto-Classification Options window providing further options to customise and manage individual classifiers.

It is also possible to translate existing classifiers into new classifiers to create more complex classification criteria in the following ways:

- **Classifier Negation** — Create a classifier that tests the opposite of a given classifier by right-clicking on the entry in the **Available Classifiers** table, selecting **Negate**, completing the **Classifier name** and **Classification type** fields and clicking the **Create** button.
- **Classifier Duplication** — Create a duplicate of an existing classifier, perhaps in order to make only a slight modification to an attribute, by selecting the entry in the **Available**

Classifiers table and clicking on the **Duplicate** button. The new copy of the classifier will appear at the bottom of the table with the word “Copy” added to its name.

- **Classifier Combination** — Combine 2 classifiers to create a new classifier that checks whether either or both criteria are satisfied by a CNV by selecting exactly 2 classifiers from the **Available Classifiers** table, right-clicking on the table and selecting the **Combine** button. In the **Advanced Classifier** window that appears, enter a **Classifier name**, select the type of combination, choose the **Classification type** which will be assigned to any CNVs that satisfy the criteria and click **Create.**

The **Advanced Auto-Classification Options** window also allows users to **Edit** or **Delete** existing classifiers via the appropriate buttons.

Resolving conflicts

If the user has chosen to apply all classifiers and to also automatically transfer the automatic classification to the initial classification, resulting in the attempted assignment of two or more different classifications, a window will appear prompting the user to resolve this conflict by selecting a single initial classification for each CNV affected (Figure 52).

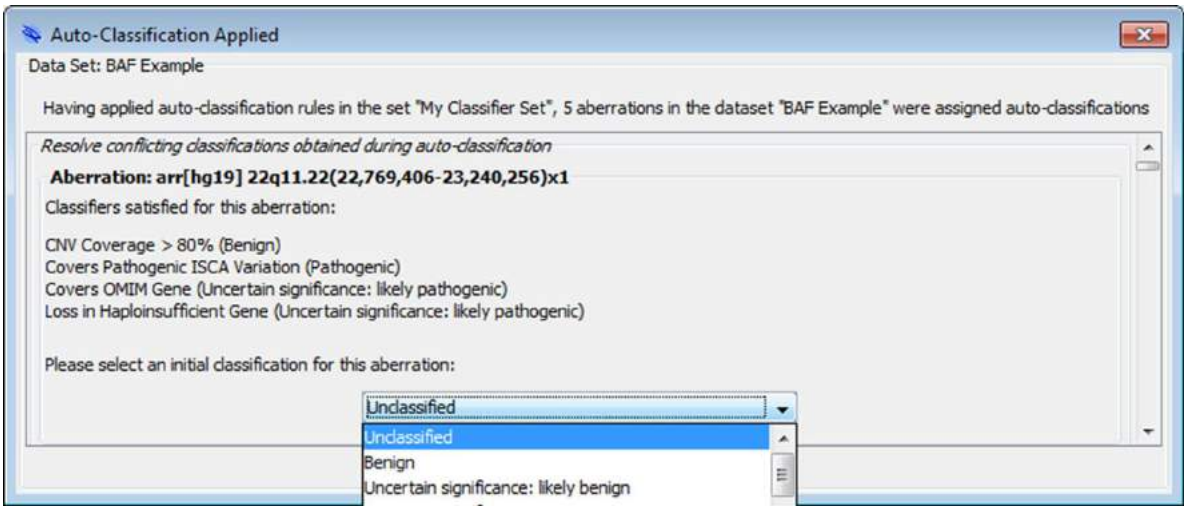


Figure 52: The interface enabling users to resolve conflicting automatic classifications by manually selecting a classification to be associated with the CNV.

Reporting and exports

Whilst it is possible to access and view all data and analyses within the CytoSure Interpret Software, it is often useful to be able to export data for use elsewhere — for example as a hard copy document or as alternative data formats which can be accessed by different software. The most common type of data export are reports, which document the results of data analysis for a particular sample in a succinct and easy to interpret document (HTML, PDF or txt). Additionally, other kinds of data export are possible from the software, such as snapshots of graphics within the software, classification information in DECIPHER format

and text file summaries of probe signal results, to name a few. These functionalities ensure that all the data can be accessed in as many ways and formats as possible, thus providing flexibility and interoperability. Refer to the following sections to discover all the ways in which data can be reported and exported using CytoSure Interpret Software.

Reports

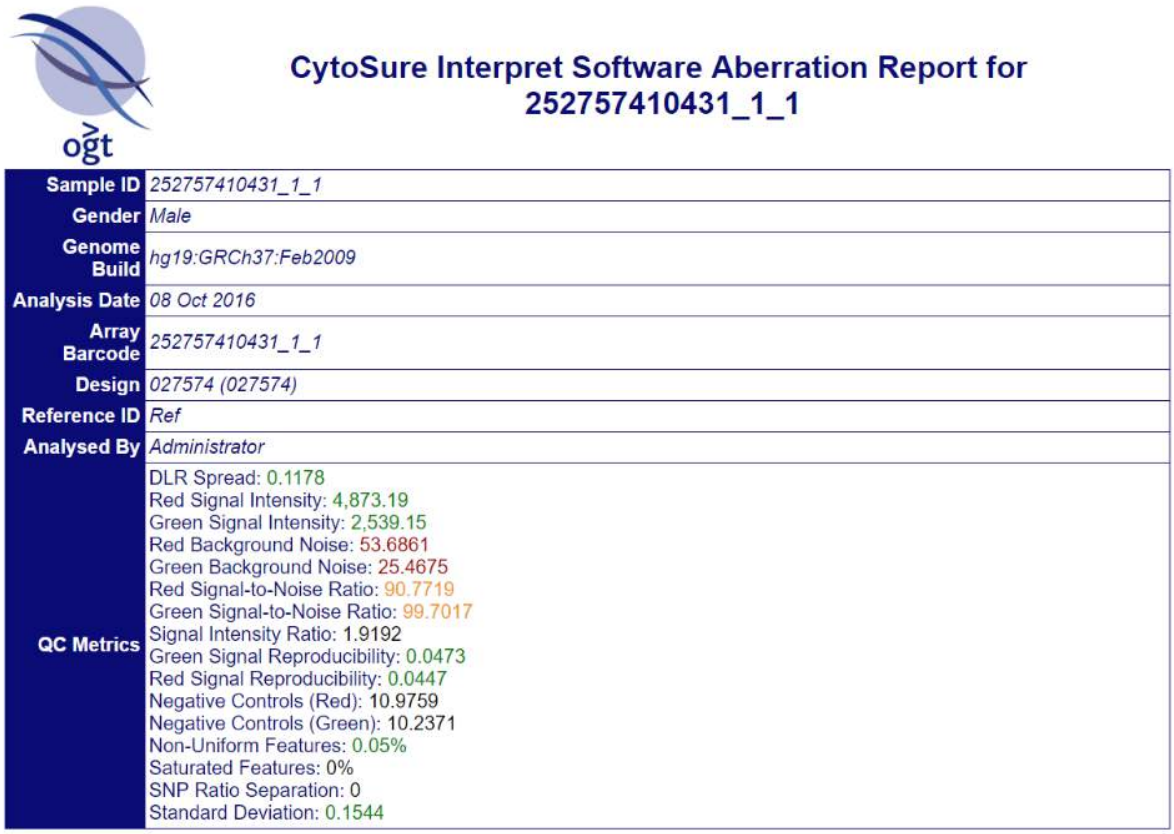
Reports are documents generated by the software to summarise the data from one sample. They provide versatility; allowing the outcome of an analysis to be viewed independently of the software and contributing to collaboration and data sharing.

By default, the reports can be generated in three different data formats: PDF, HTML and tab-delimited text. Each of these formats displays the same information, with the PDF and HTML reports additionally including images and further details of each aberration and ideograms of the chromosomes.

Standard report content

The content in the reports has a standard format with further options for customisation. The structure and content of the reports when generated using the default **OGT Protocol** is outlined below:

- **Sample data**
    - Specific fields are specified in the **Options** menu, but can include Sample ID, Date, HPO Phenotype or Expected Barcode.
    - The default configuration uses the fields: Sample ID, Gender, Phenotype and Notes.
    - See Figure 1 for an example.



CytoSure Interpret Software Aberration Report for 252757410431_1_1

Sample ID	252757410431_1_1
Gender	Male
Genome Build	hg19:GRCh37:Feb2009
Analysis Date	08 Oct 2016
Array Barcode	252757410431_1_1
Design	027574 (027574)
Reference ID	Ref
Analysed By	Administrator
QC Metrics	DLR Spread: 0.1178 Red Signal Intensity: 4,873.19 Green Signal Intensity: 2,539.15 Red Background Noise: 53.6861 Green Background Noise: 25.4675 Red Signal-to-Noise Ratio: 90.7719 Green Signal-to-Noise Ratio: 99.7017 Signal Intensity Ratio: 1.9192 Green Signal Reproducibility: 0.0473 Red Signal Reproducibility: 0.0447 Negative Controls (Red): 10.9759 Negative Controls (Green): 10.2371 Non-Uniform Features: 0.05% Saturated Features: 0% SNP Ratio Separation: 0 Standard Deviation: 0.1544

Figure 1: Top section of HTML and PDF report showing the sample data, experiment details and QC metrics fields.

- **Experiment details**
  - Specific fields are specified in the **Options** menu, but can include FE file, Project, Analysis Protocol or Reference ID.
  - The default configuration uses the fields: Analysis Date, Array Barcode, Design, Reference ID and Analysed By.
  - See Figure 1 for an example.

- **QC metrics**
  - By default, all quality control metrics are displayed. It is possible to choose subsets to display by navigating to the **QC Metrics** tab of the **Options** menu.
  - The metrics in the PDF and HTML reports are displayed in the traffic light colour scheme. This scheme is also used elsewhere in the software to indicate the quality

of the metric value (Excellent, Satisfactory or Poor). See Quality Metric Review for more information on quality metrics.

- See Figure 1 for an example.

- **Ideogram of all chromosomes**
  - The ideogram is present in the PDF and HTML reports but not the text format.
  - The ideogram shows a graphic of each chromosome with red and green/blue* blocks to indicate loss and gain CNVs respectively. If the data includes SNPs, the CNVs will be displayed on the right and the SNPs will be displayed on the left, with the level of homozygosity across the genome found above. For help interpreting this graphic, see the CNVs/SNPs tab section.

    *Gains are shown in green by default but can be changed to blue, a setting which is found in the **Display** tab of the **Options** menu.
  - These info-graphics match the ideogram image in the **Detected CNVs/SNPs** section of the **CNVs/SNPs** tab (whilst on the **Ideogram View**).
  - See Figure 2 for an example.

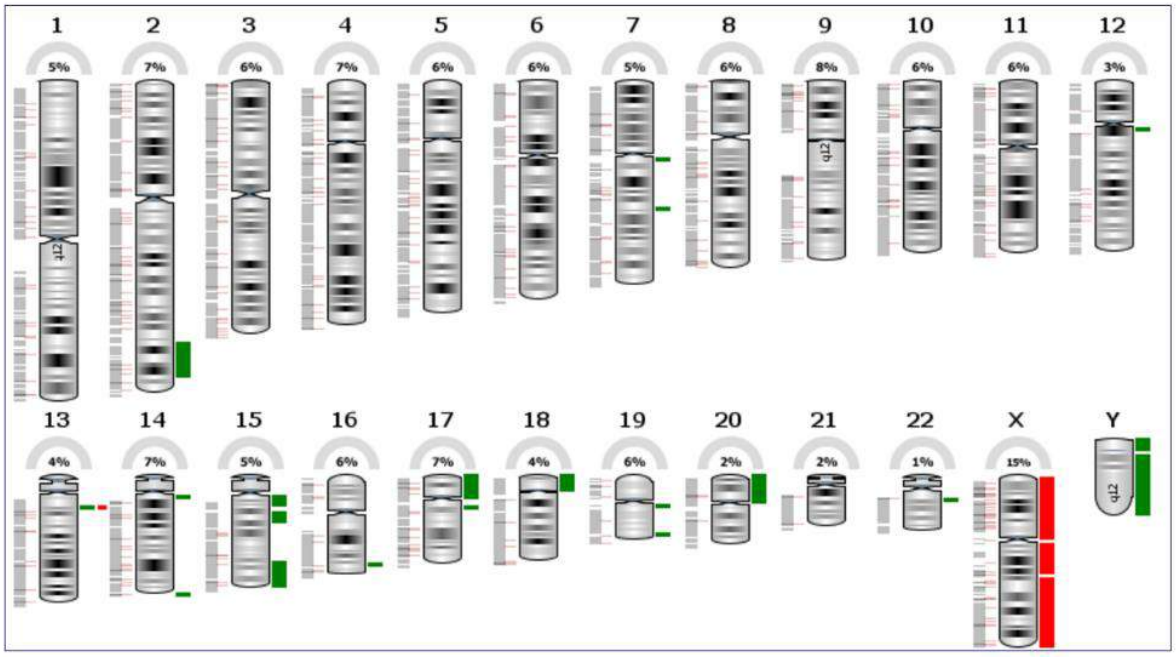


Figure 2: Ideogram of all chromosomes as displayed in the HTML and PDF reports.

- **Table of detected CNVs**
  - Table detailing all the filtered CNVs.

- Present in all report types.
- The columns in the tables are as follows: Chromosome, Cytogenetic Location, Start, Stop, Size, P-Value and Gain/Loss.
- For an example see Figure 3.

	Chromosome	Cytogenetic Location	Start	Stop	Size	P-Value	Gain/Loss
1	2	q33.2q35	204598942	219053447	14.45Mb	0	Gain
2	2	q35q37.1	219992599	231288968	11.3Mb	0	Gain
3	7	q11.1q11.21	61081314	62353488	1.27Mb	0.6778	Gain
4	7	q22.1	98789198	98856948	67.75Kb	0.5044	Gain
5	12	q11q12	37896066	38276243	380.18Kb	0.3605	Gain
6	13	q12.2	28608193	28608262	69b	0.9384	Gain
7	13	q12.2	28608340	28608436	96b	0.8736	Loss
8	14	q11.2	19457944	20420346	962.4Kb	0.2624	Gain
9	14	q32.33	106057624	107286764	1.23Mb	0.5174	Gain
10	15	q11.1q13.1	20003669	28704700	8.7Mb	0.0048	Gain
11	15	q14q15.1	34533380	42636858	8.1Mb	0	Gain
12	15	q14q15.2	34533380	43438540	8.91Mb	0.0022	Gain
13	15	q25.1q26.3	78845710	102399814	23.55Mb	0	Gain
14	16	q23.2	79631172	79633727	2.56Kb	0.3	Gain
15	17	p13.3p11.2	10152	21416908	21.41Mb	0	Gain
16	17	p13.3p11.2	10152	22198674	22.19Mb	0.0002	Gain

Figure 3: Table of detected CNVs displayed in HTML and PDF reports.

- **Table of chromosome homozygosity**
  - Table detailing the percentage of homozygosity observed for each chromosome (in descending order).
  - Present in all report types as long as the input data includes SNPs.
  - The columns in the tables are as follows: Chromosome, Number of SNPs, Number of homozygous SNPs, Number of heterozygous SNPs, Number of SNPs with unknown genotype, Percentage homozygosity.
  - For an example see Figure 4.

Chromosome	# SNPs	Homozygous	Heterozygous	Unknown Genotype	% Homozygosity
X	349	52	288	9	15%
9	251	19	229	3	8%
14	208	15	187	6	7%
17	205	14	187	4	7%
2	584	38	530	16	7%
4	417	27	383	7	7%
11	330	21	304	5	6%
3	514	32	468	14	6%
6	401	25	366	10	6%
5	390	24	354	12	6%
10	343	21	315	7	6%
19	97	6	91	0	6%
8	380	23	352	5	6%
16	180	10	167	3	6%
1	472	24	435	13	5%

Figure 4: Table of chromosome homozygosity displayed in HTML and PDF reports.

- **Table of LOH regions**
  - Table detailing the identified LOH regions.
  - Present in all report types as long as the input data includes SNPs.
  - The columns in the tables are as follows: ISCN notation (International System for Human Cytogenetic Nomenclature), LOH region size, Number of SNPs within region and LOH region score.
  - For an example see Figure 5.

LOH Regions			
ISCN	Size	# SNPs	Score
arr[hg19] Xp21.1p11.3(34,170,560-43,278,161)x2 hmz	9.11 Mb	23	10

Figure 5: Table of LOH regions displayed in HTML and PDF reports.

- **Aberration details**

- The final section of the reports lists all of the filtered CNVs with some associated information.
- In text reports, this information is displayed in a tab delimited table like that seen in Figure 6. Table columns are as follows: ISCN notation, Gain/loss status, Number of probes within aberration, Mean log ratio, Size of aberration, Initial classification and Final classification.
- In the HTML and PDF reports, the aberrations are presented in a more visual manner, as seen in Figure 7. All the information given in the text reports is displayed below a visual image (snapshot) of the aberration as shown in the **Genomic View** tab of the software.

Aberrations

ISCN Notation	Gain/Loss	# Probes	Mean Log Ratio	Size	Classification (Initial)	Classification (Final)
arr[hg19] 2q33.2q35(204,598,942-219,263,558)x3	Gain	92	0.6312	14.66Mb	Unclassified	Pathogenic
arr[hg19] 2q35q37.1(219,992,599-231,235,872)x3	Gain	71	0.6458	11.24Mb	Unclassified	Unclassified
arr[hg19] 7q11.1q11.21(61,081,314-62,353,488)x3	Gain	9	0.1891	1.27Mb	Benign	Unclassified

Figure 6: Tab delimited table of aberrations displayed in text reports.

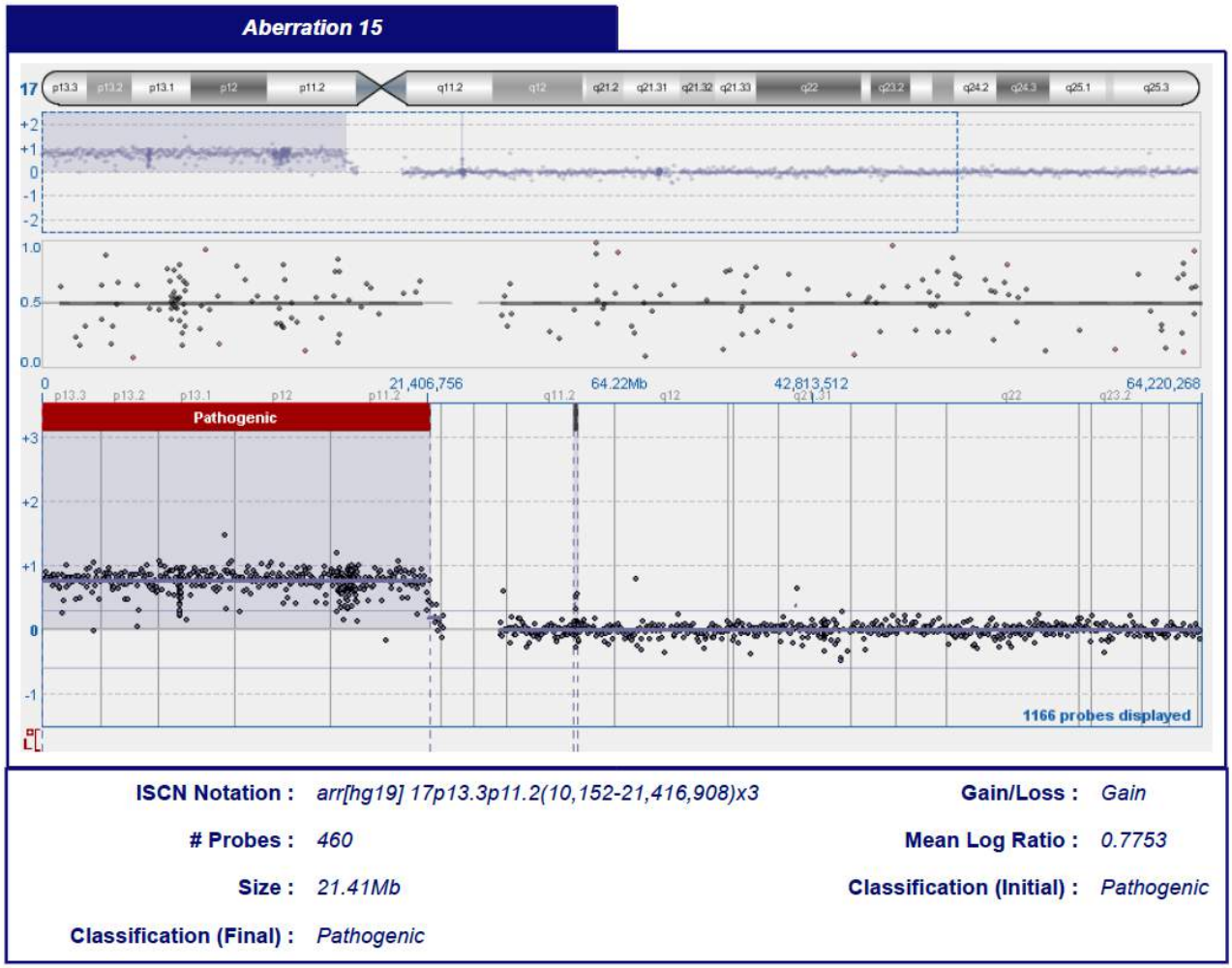


Figure 7: Visual and information display of an aberration as shown in HTML and PDF reports.

Customisable report content

Annotation reporting

Basic annotation reporting

Aberration reports can be configured to include the names of any annotations overlapping detected CNV regions on a track-by-track basis. This will add an additional field to the **Aberration Details** boxes of the PDF and HTML reports. This can be seen in Figure 8 where the **Genes** field specifies the gene which overlaps the CNV from the default **Gene** track (taken from Ensembl).

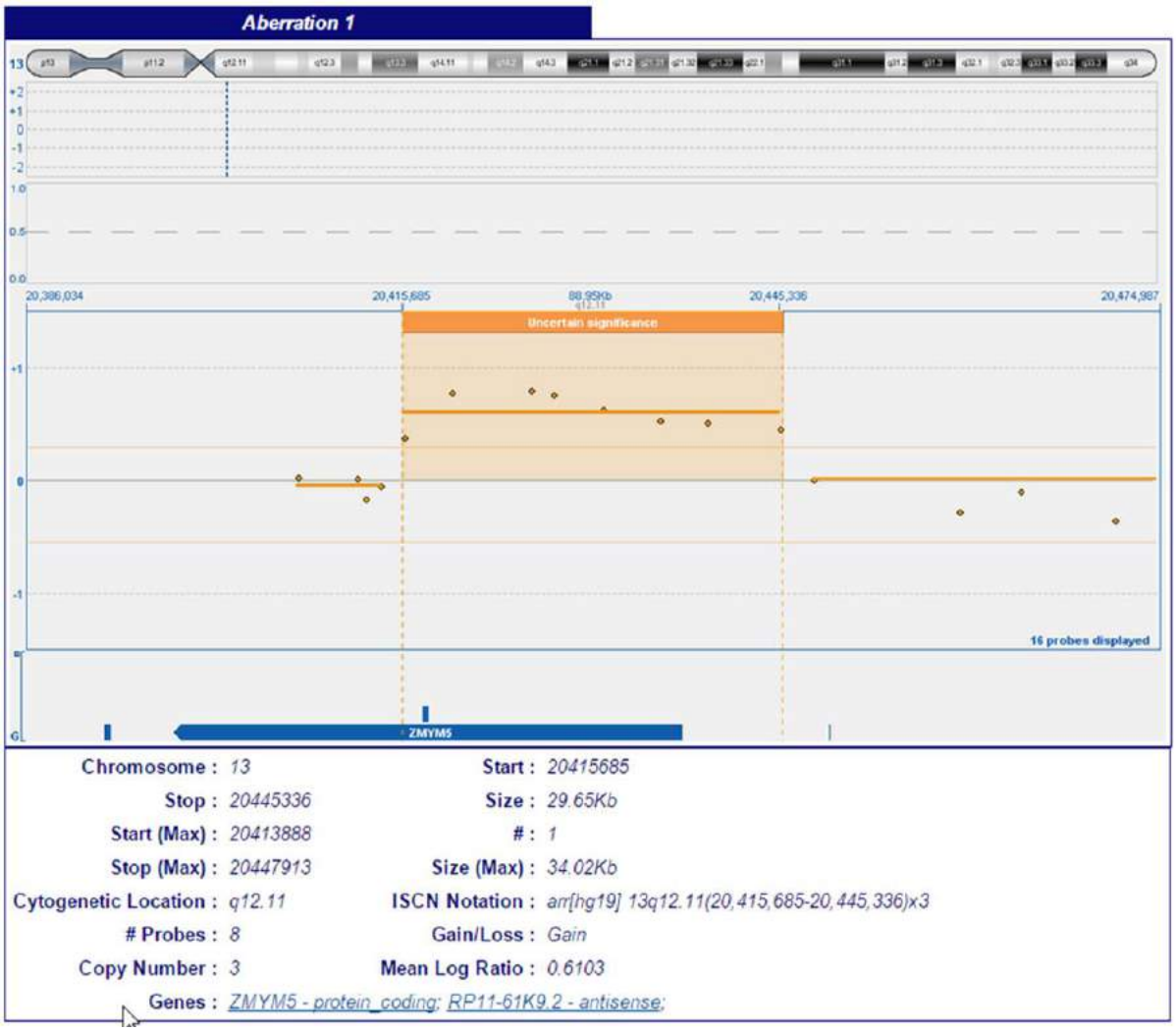


Figure 8: Genes field representing overlapping annotation information. Present in the Aberration Details section of HTML and PDF reports.

In order to add annotation track details to the report:

1. Navigate to **Tools -> Options... -> Annotation** and choose the **CNV** or **LOH Reporting** tab on the table. The resulting window can be seen in Figure 9, where the table of interest is in the lower portion of the window.

2. All of the possible aberration tracks are listed in rows with tick boxes for each labelled **Report, Report #** and **Report %**. Select the relevant **Report** tick box to include the names of annotations from this track in the report. The **Report #** tick box will report the number of annotations covering the aberration and the **Report %** tick box will report the percentage of the aberration covered by annotations.

3. Navigate to the **Annotation Info** section of the **Report Generation** tab of the **Options...** menu.

4. Add the required fields to the **Included** box, as described in General options.

5. Generate the report as described in Report generation.

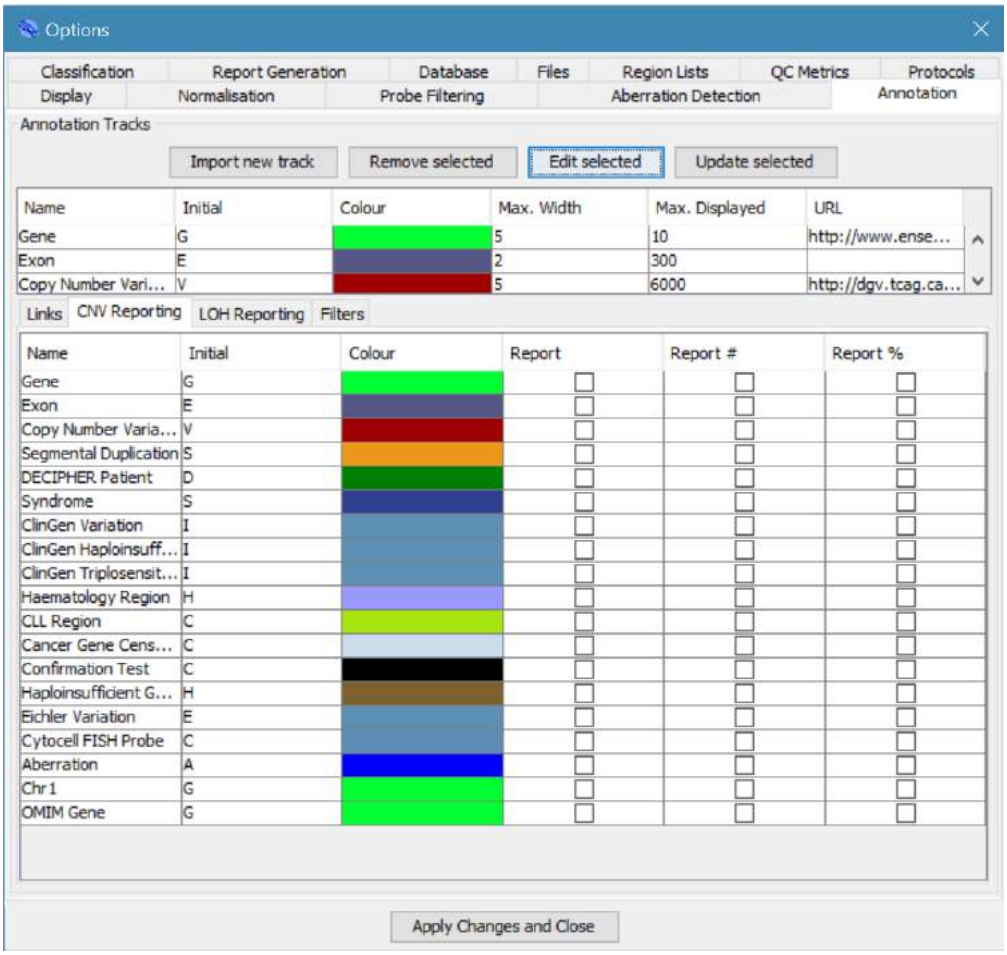


Figure 9: Annotation tab of Options menu.

Custom annotation reporting

It is also possible to customise the text in the report to include other information related to the annotation as well as, or instead of, its displayed name. For example, as all annotations in the default **Gene** track (taken from Ensembl) include a “Gene Biotype” custom field (all fields can be displayed by hovering over an annotation [see Figure 10]), CytoSure Interpret Software can be configured so that both pieces of information are included when the **Genes** field is part of the report.

Figure 10: Gene information available for inclusion in reports.

To do this:

1. Enable reporting of the annotation name by selecting **Tools -> Options… -> Annotation -> CNV Reporting** and ensure that the **Report** column of the annotation track is checked.
2. Right-click in the same cell of the table as the checkbox and select **Advanced Options** (Figure 4).

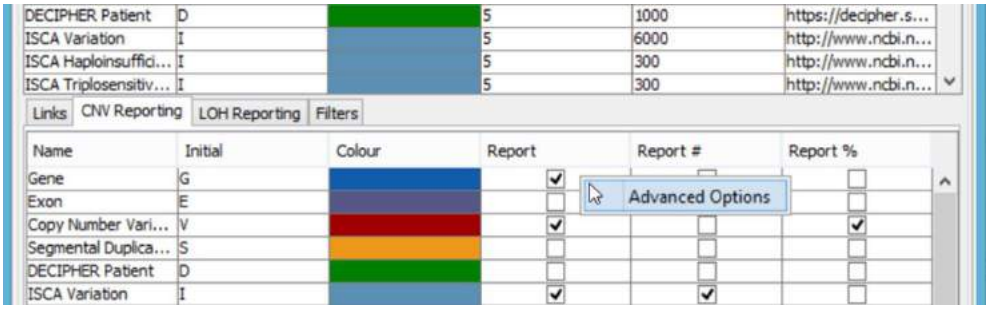


Figure 11: Accessing advanced CNV Reporting options for annotation tracks.

3. In the **Customise annotation information** section of the **Advanced annotation options** window (Figure 12), enter the format to be used to report overlapping annotations in the selected track using the same dynamic field codes described in URL Composition. For example, to report the name of the annotation and the value of the "Gene Biotype" field, separated by a hyphen, set this value to "{#name} – {Gene Biotype}".

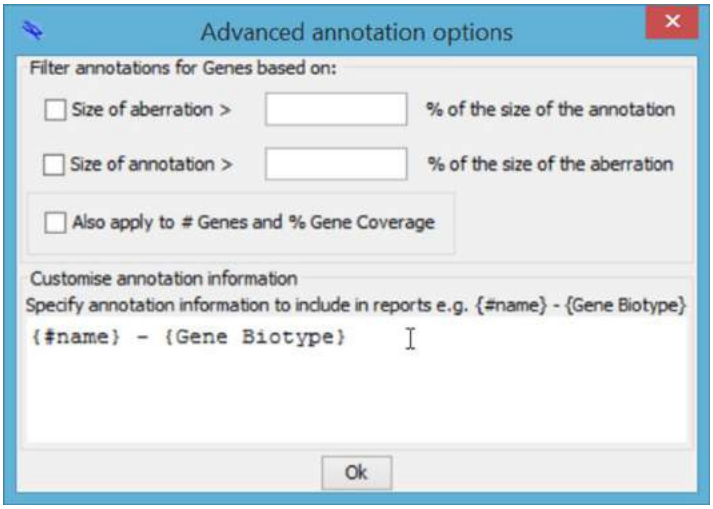


Figure 12: Setting the annotation reporting format.

When reports are subsequently generated that have been configured to include the names of overlapping annotations in the track (as described in Basic annotation reporting), they should now be displayed in the specified format.

Probe-annotation overlap reporting

The **Table View** tab contains information for all probes currently displayed in the **Genomic View**, including probe name, location, signal intensities and GC content. This table can be customised to also include the names of any annotations which cover the probe (e.g. genes). Configuration can be carried out as follows:

1. Enable reporting of the annotation name by selecting **Tools -> Options… -> Annotation -> CNV Reporting** and ensure that the **Report** column of the required annotation track is checked.
2. In the **Table View** tab, right-click on the table header and select the required column from the pop-up menu displayed (Figure 13).

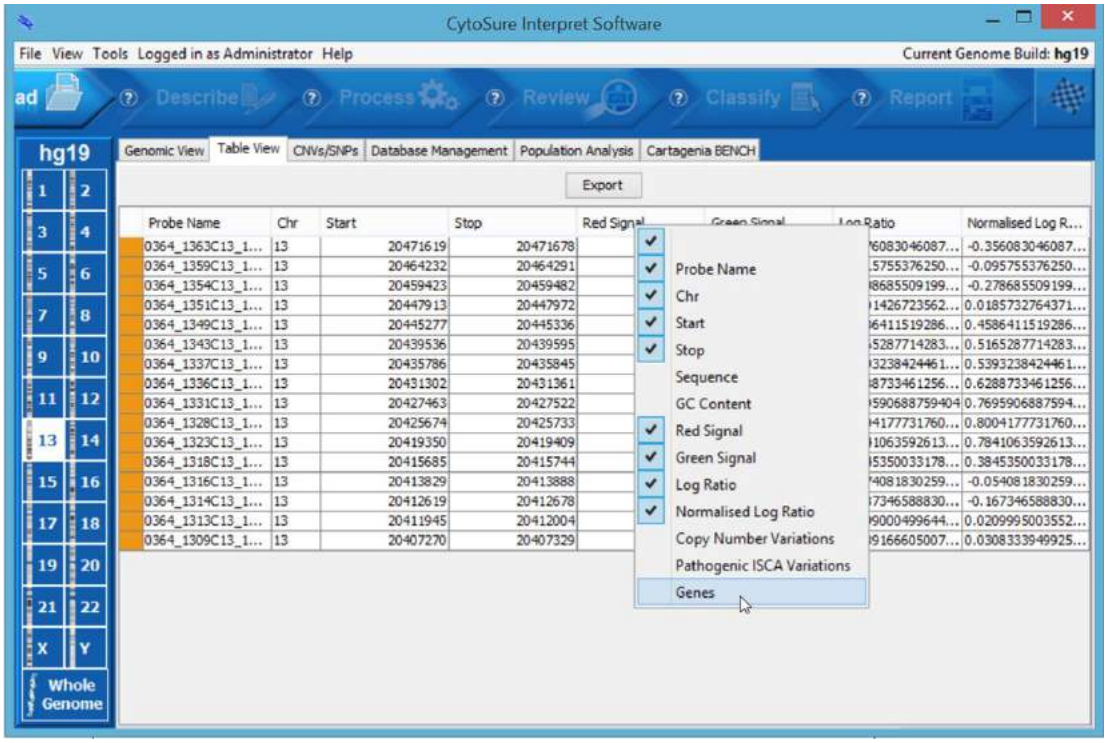


Figure 13: Adding annotation overlap information columns to the Table View.

If a large number of probes are displayed in the **Genomic View**, a large number of calculations will need to be made to determine which annotations overlap with the probes in the table. To reduce the time for these calculations to be made, load the required annotation tracks into memory by right-clicking on the track in the **Genomic View** and selecting **Load into memory**.

Custom annotation information display

The Aberration annotation track contains CNVs detected in previously analysed cases, and hovering over a CNV in the track results in the display of a number of pieces of information relating to the selected CNV. This content is called a **Tooltip**. Customisation can be carried out by right-clicking within the Aberration track and using the **Tooltip Fields** menu to individually select the information that should be displayed when the user hovers the mouse over an annotation in this track (see Figure 14).

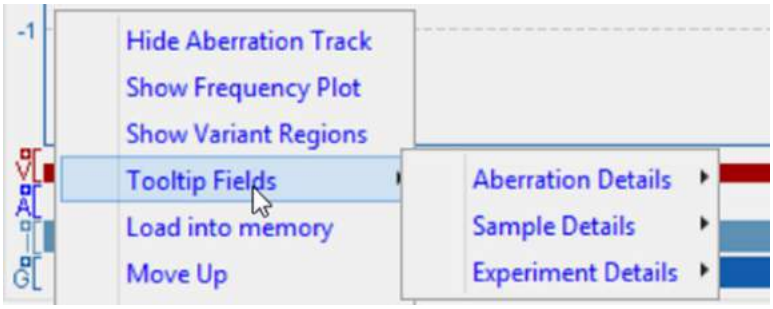


Figure 14: Aberration track tooltip customisation.

Fields are grouped into three categories — **Aberration Details**, **Sample Details** and **Experiment Details** — and users are able to select a number of fields for inclusion in the tooltip to make the view as useful as possible within their analysis workflow.

Classification reporting

As with annotation reporting, there are also options to customise the reporting of aberration classifications. These options can be found in the **Classification** tab of the **Options** menu (Figure 15 and Figure 16).

It possible to specify which aberrations to report on the basis of their classification. This could be useful if, for example, the user does not want benign CNVs included in the report. Click on **Tools** -> **Options** -> **Classification** and deselect the **Report?** tick box to exclude aberrations classified with the selected classification from any generated reports (Figure 15). Please note that this does not affect the submission of aberrations to the database.

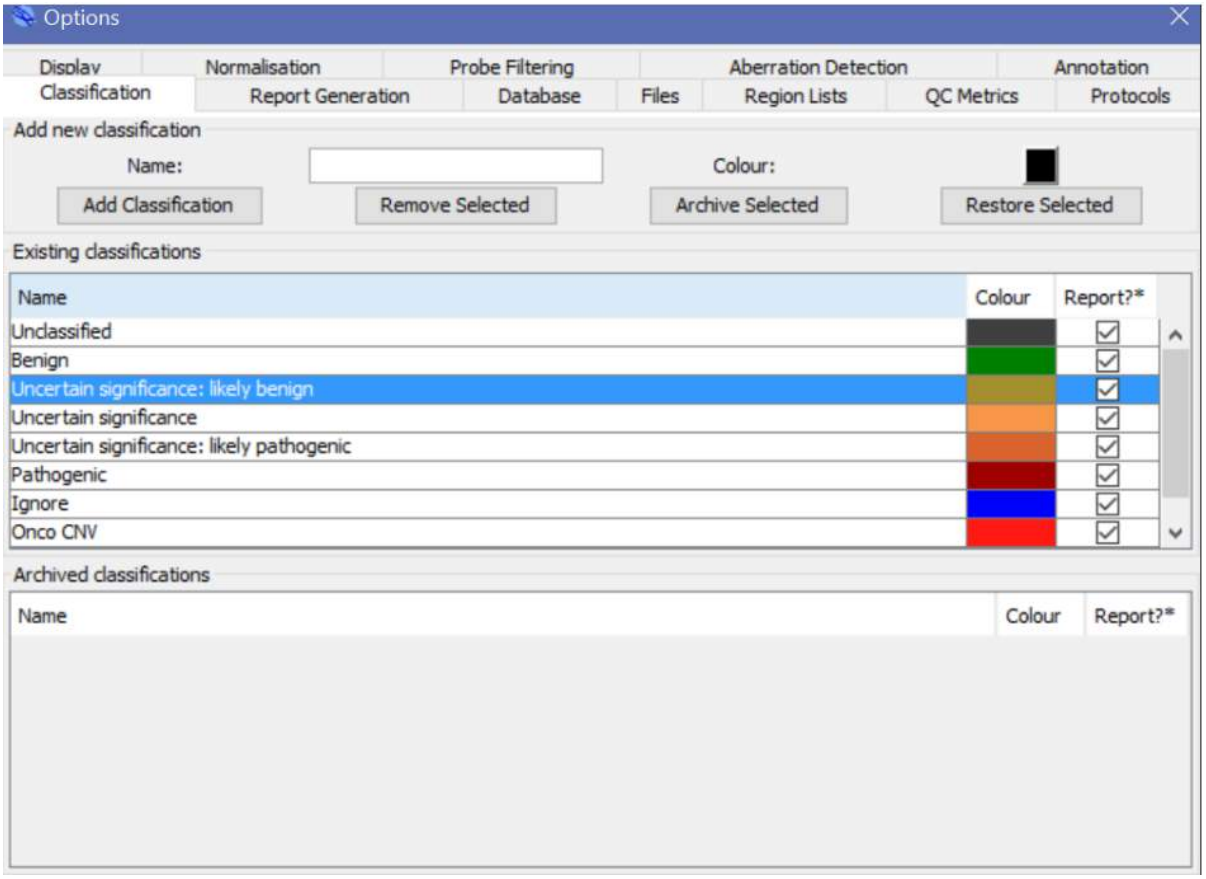


Figure 15: Upper section of Classification tab of Options menu.

By default, the aberrations excluded by deselecting the **Report?** options will still be included in the summary table at the beginning of the report and the experiment ISCN notation. However, it is possible to remove these aberrations by selecting the **Also remove these aberrations from the summary table?** check box and/or the **Also remove these aberrations from the experiment ISCN Notation?** checkbox, both located in the ***Reported Classifications** section (Figure 16).

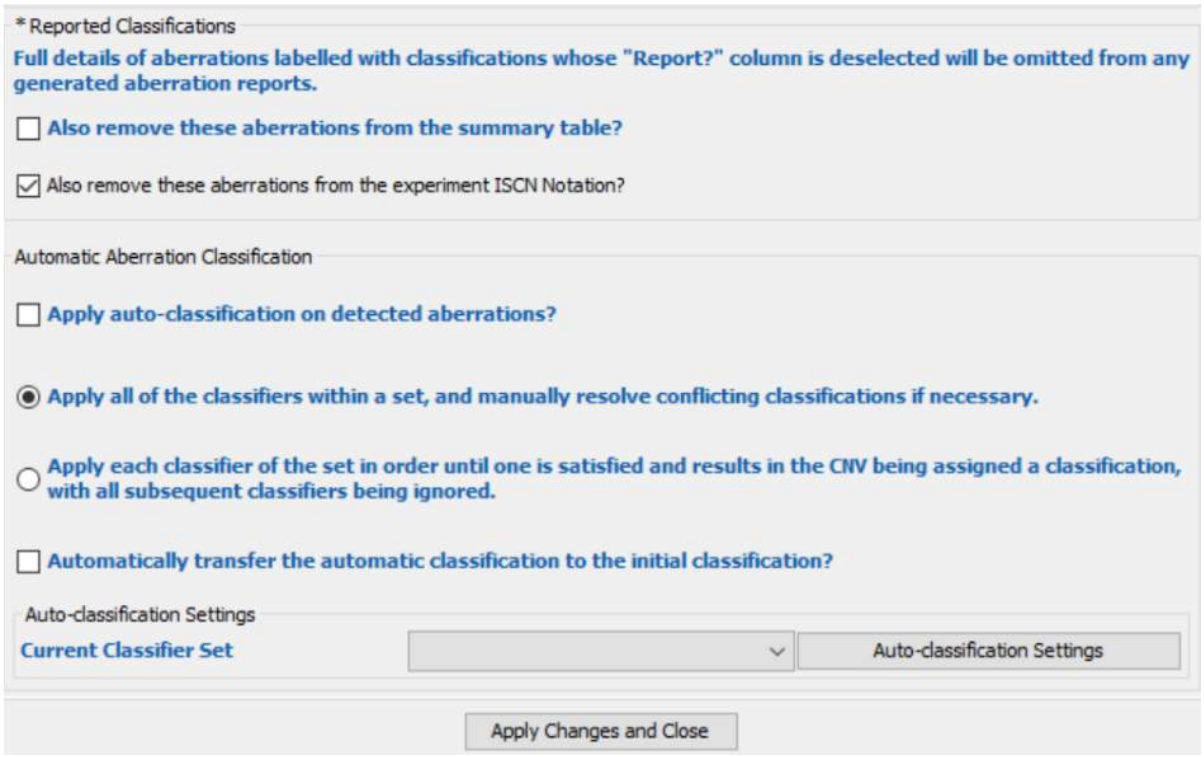


Figure 16: Lower section of Classification tab of Options menu.

Other options

General options

The **Report Generation** tab of the **Options** menu contains a number of general options for the reports (**Tools -> Options... -> Report Generation**). As seen in Figure 17, there are two sections of this window which refer to report content: **Logo** and **Report Content**.

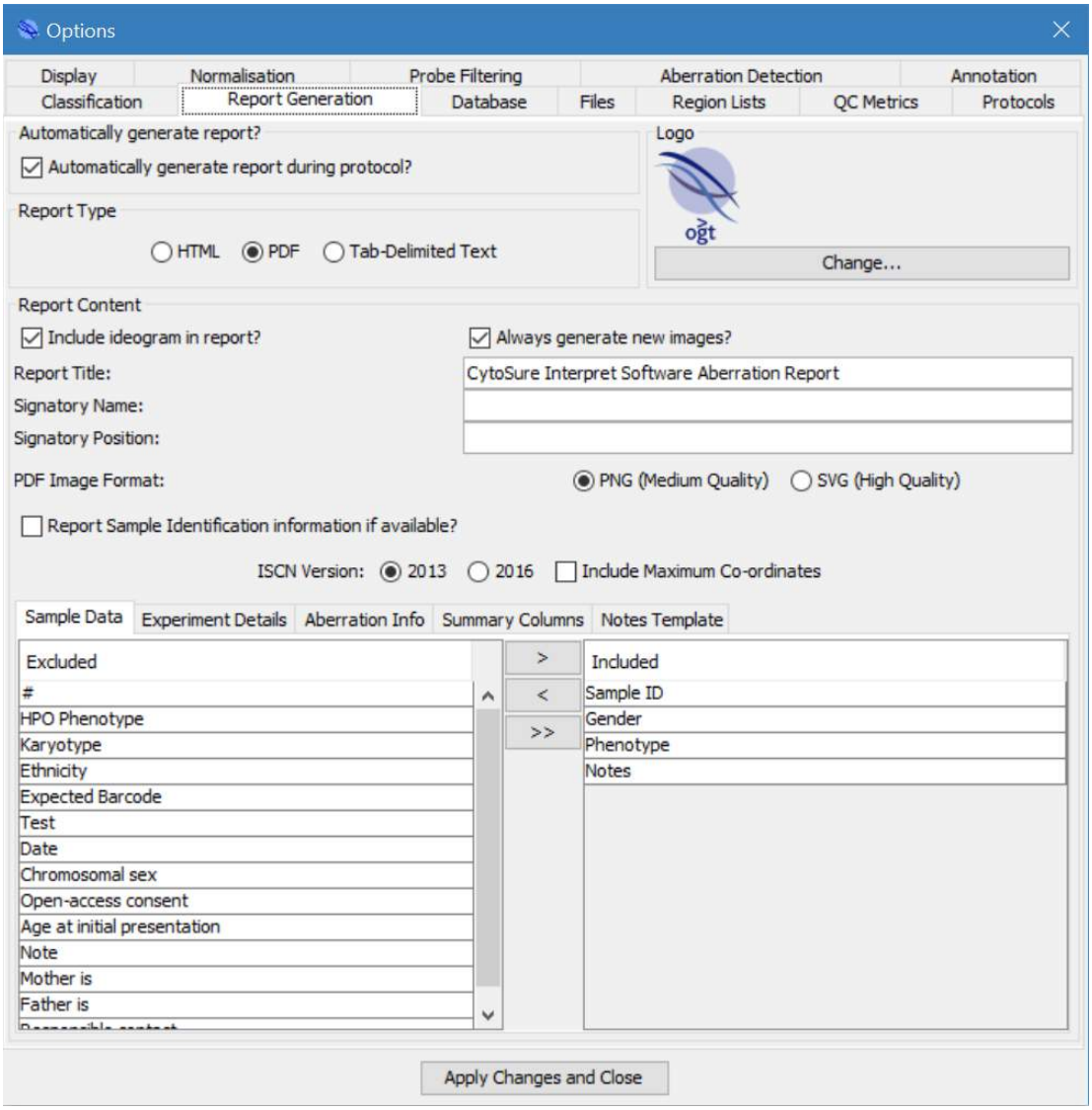


Figure 17: Report Generation tab of Options menu.

The **Logo** section (Figure 17, top right) allows the user to select a logo to display at the top of the front page of PDF and HTML reports. The **Report Content** section (Figure 17, lower half), contains several further options:

- **Include ideogram in report? —** This option determines whether the ideogram of chromosomes is displayed at the bottom of the first page (in PDF and HTML report). This ideogram is a copy of the one seen in the software on the **SNPs/CNVs** tab when on the **Ideogram View.** An example of it in the report is shown in Figure 18.

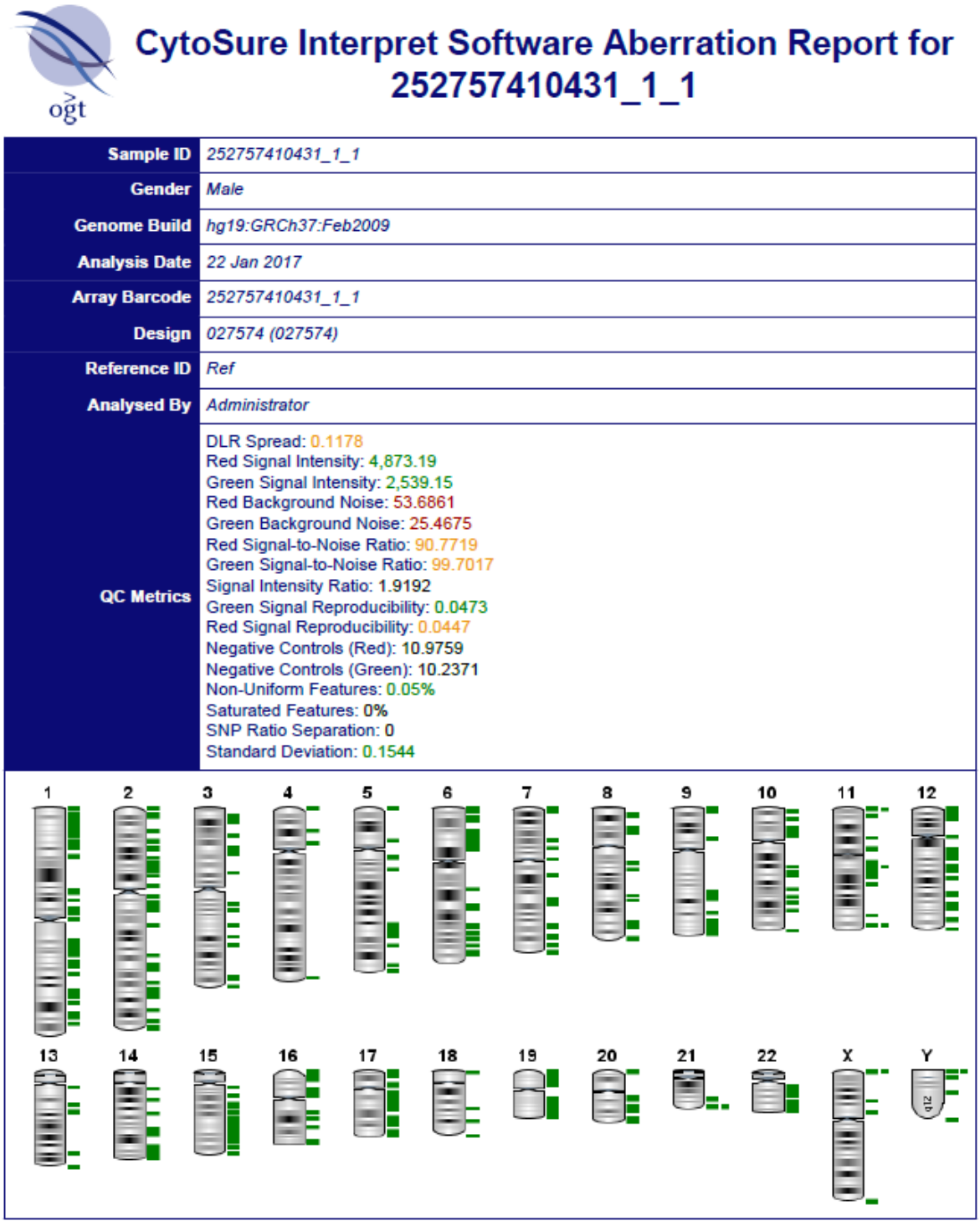
- **Always generate new images? —** When a PDF or HTML report is first generated, each of the CNVs are imaged and saved in a folder labelled **filename.[pdf/htm]_images** in the same folder as the report. These images are used in the **Aberration Details** panels, as seen in Figure 19. If this tick box is not checked and a change is made which alters the image (e.g. hiding datapoints, or classification of the aberration), the images will not be re-captured and re-generated reports will use the previously captured images. This also means that if new CNVs are specified, whilst they will be added to the **Aberrations Table,** they will not have an **Aberration Details** panel in the report. Therefore, it is advised that this tick box remains checked.

- **Report title**

- **Signatory Name** and **Position —** If specified, these details would be placed at the base of the final page of PDF and HTML reports.

- **Report Sample Identification information if available? —** If the microarray design includes spike-in probes for tracking, then selecting this option will result in the inclusion of a pooled summary of the spike-in probes in the report.
- **ISCN Version —** ISCN is a system used to describe cytogenetic rearrangements. In the reports, the notation is used in the **Aberration Details** panels (Figure 19). The nomenclature varies slightly between the 2013 and the 2016 specifications and this option allows the user to choose which they prefer to use. Selection of the **Include Maximum Co-ordinates** option is used to specify the largest possible region, from the end of the last probe before aberration to the beginning of the first probe after the aberration.

- **Data excluded and included —** The lower section of the page is used to specify which details to include in different sections of the report. The **Sample Data** and **Experiment Details** tabs are used to specify which information to include in the section at the beginning of the reports (Figure 17). The **Aberration Information** tab is for information in the **Aberration Details** panels of the report (Figure 19) and the **Summary Columns** tab for columns in the **Table of Detected CNVs** in the reports.



CytoSure Interpret Software Aberration Report for 252757410431_1_1

Sample ID	*252757410431_1_1*
Gender	*Male*
Genome Build	*hg19:GRCh37:Feb2009*
Analysis Date	*22 Jan 2017*
Array Barcode	*252757410431_1_1*
Design	*027574 (027574)*
Reference ID	*Ref*
Analysed By	*Administrator*
QC Metrics	DLR Spread: 0.1178 Red Signal Intensity: 4,873.19 Green Signal Intensity: 2,539.15 Red Background Noise: 53.6861 Green Background Noise: 25.4675 Red Signal-to-Noise Ratio: 90.7719 Green Signal-to-Noise Ratio: 99.7017 Signal Intensity Ratio: 1.9192 Green Signal Reproducibility: 0.0473 Red Signal Reproducibility: 0.0447 Negative Controls (Red): 10.9759 Negative Controls (Green): 10.2371 Non-Uniform Features: 0.05% Saturated Features: 0% SNP Ratio Separation: 0 Standard Deviation: 0.1544

Figure 18: Front page of PDF/HTML report including the ideogram at the bottom.

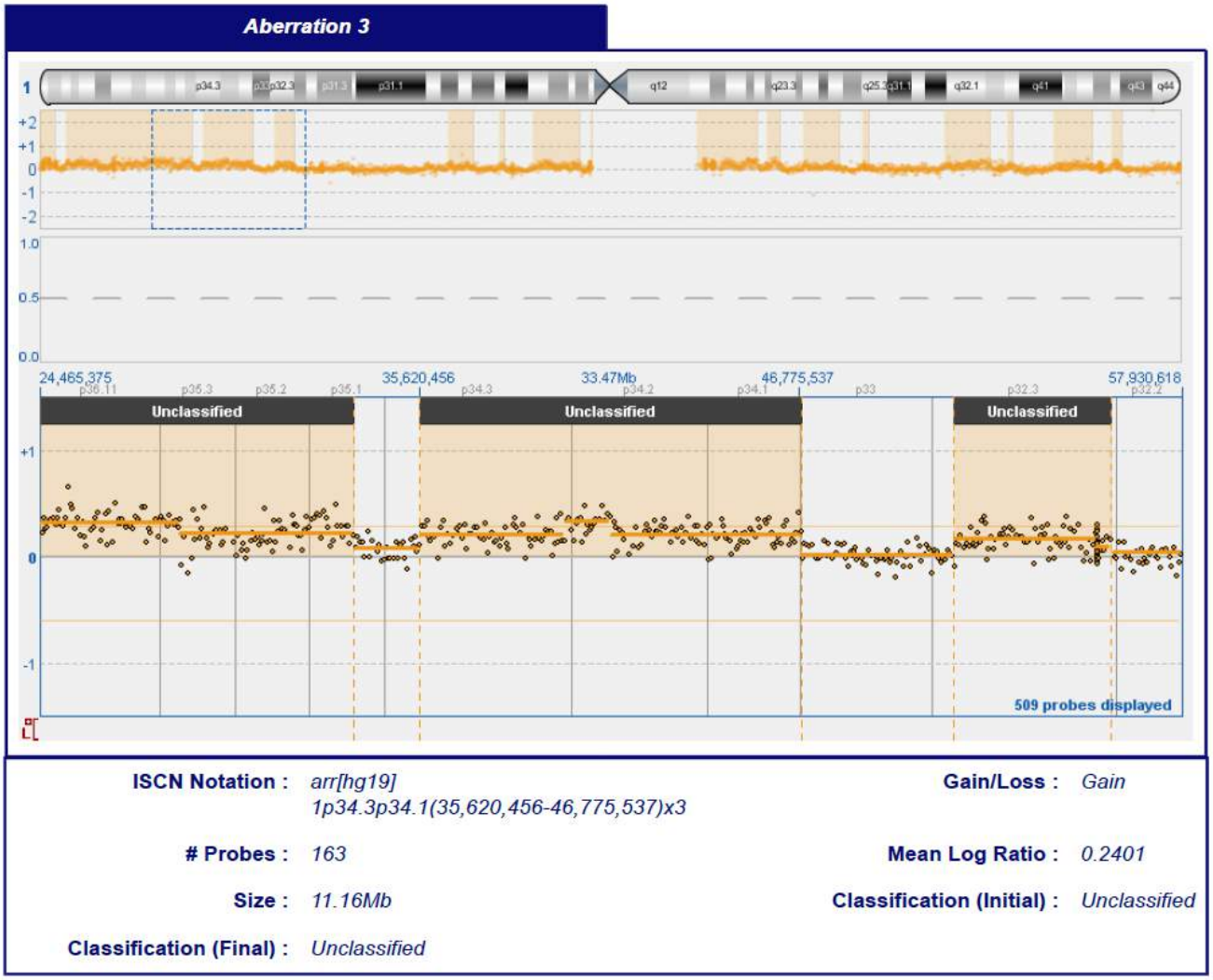


Figure 19: Example Aberration Details panel from a PDF/HTML report.

Display

The **Display** tab of the **Options** menu, as seen in Figure 20, contains settings used to edit the display of images and graphs within the software (**Tools -> Options... -> Display**). As the images in the PDF and HTML reports are directly captured from the software, these settings also affect the report display. Potentially the most useful of these options is the **Colour Scheme**, which is used across the software to indicate different samples. In the HTML and PDF reports, this colour scheme is used in the **Aberration Details** panels, for example in Figure 19 the colour is orange. In the **Options** menu, **Colour 1** refers to the colour assigned to the first loaded sample, **Colour 2** to the second loaded sample and so on.

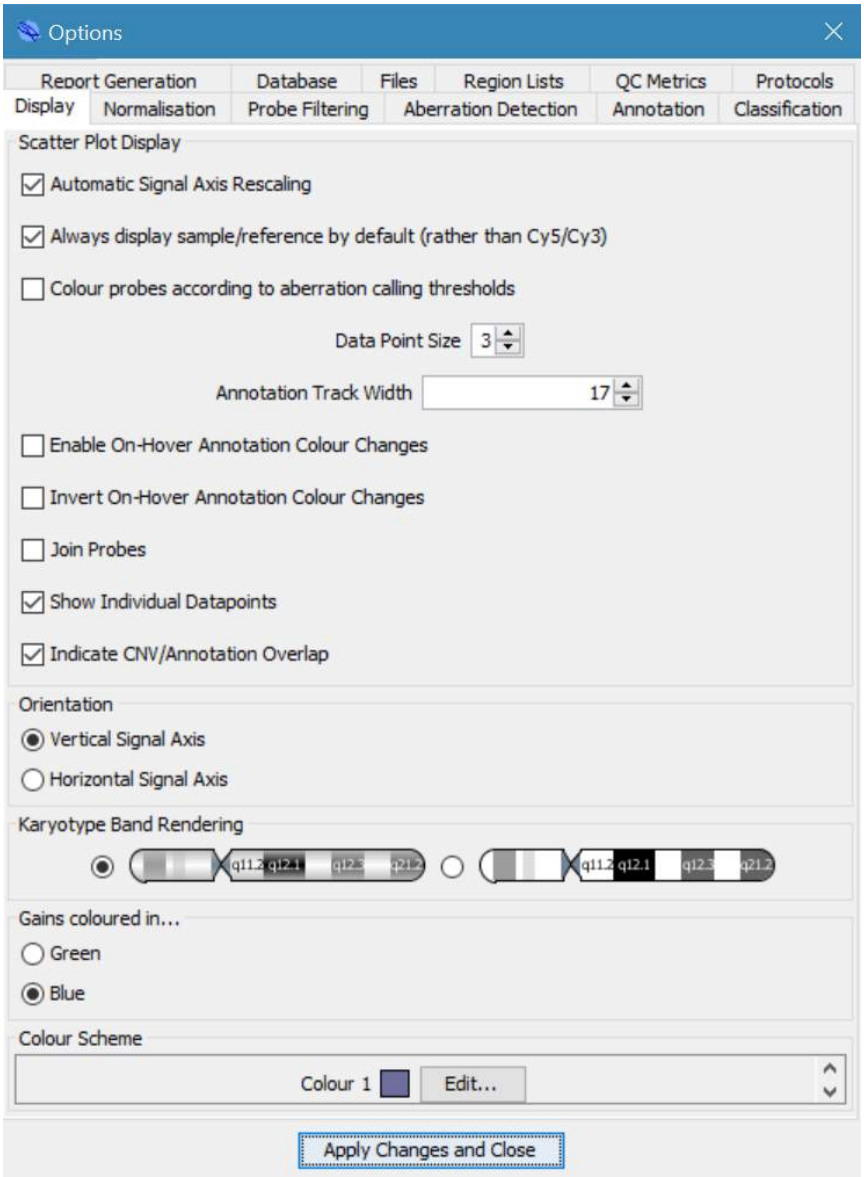


Figure 20: Display tab of Options menu.

Quality control metrics

Navigating to **Tools -> Options... -> QC Metrics** will open the QC Metrics settings (Figure 21). From this menu, it is easy to select which metrics are displayed in the report, simply by selecting (or deselecting) the boxes in the right-hand most column of the **Thresholds** table. The Quality metric review section contains further details.

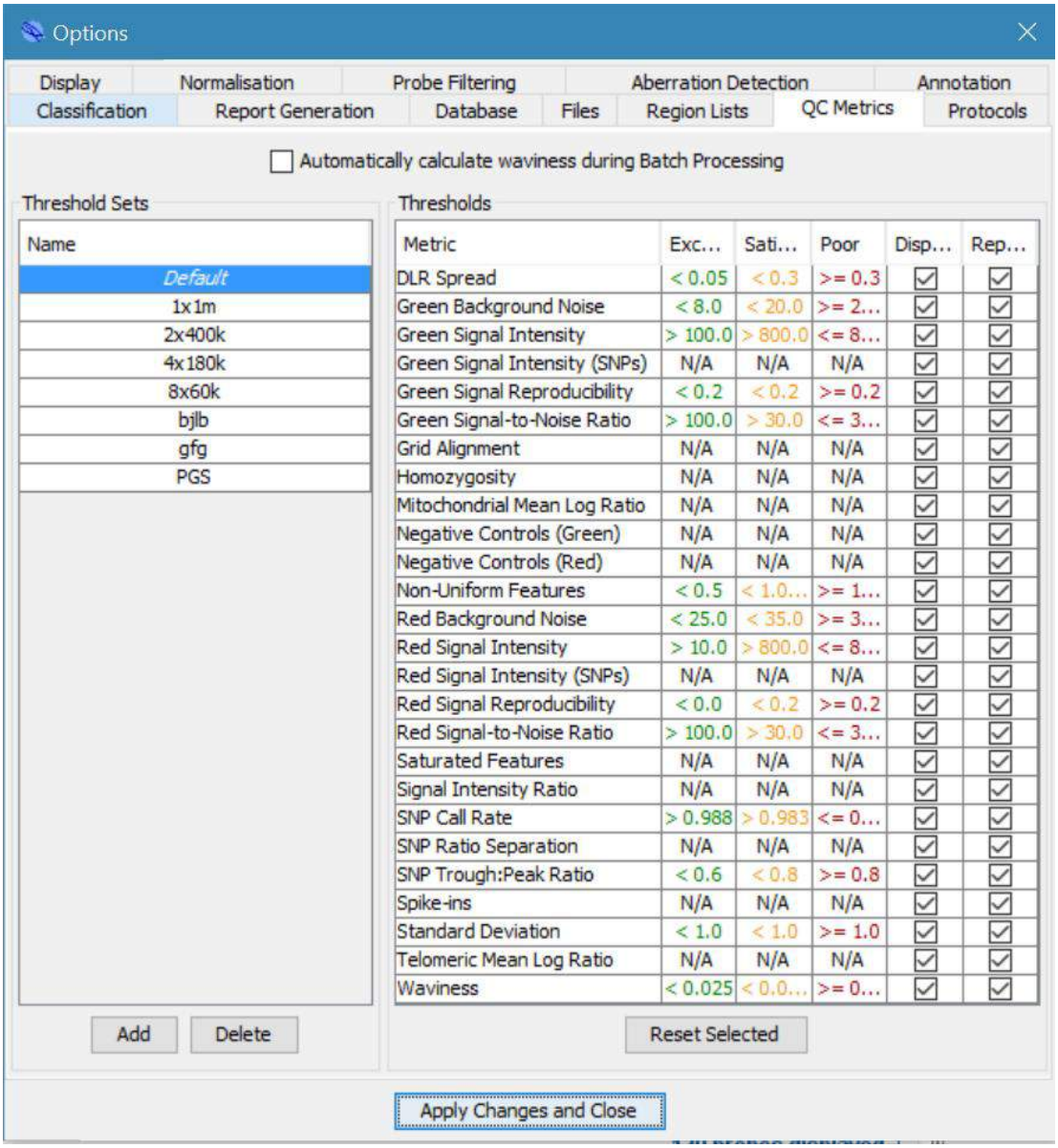


Figure 21: QC Metrics tab of Options menu.

Report generation

The final stage of the process is generating a report. There are several different ways to do this, as outlined below.

Using Accelerate Workflow Bar

The accelerate workflow bar (Figure 22) highlights the workflow for data analysis, simplifying the process. To start an analysis using the workflow, click on the **Load** button, followed by the other buttons when appropriate. When the data is ready to be reported, press the **Report** button to automatically generate a report.

- To use the buttons in the workflow, it is necessary to have a protocol loaded. This can be done through the **Protocols** menu: **Tools -> Options -> Protocols**.
- The report is only automatically generated using the **Report** menu button if the **Automatically generate report during protocol** option is selected in the **Tools -> Options -> Report Generation** tab**.**

Figure 22: Accelerate Workflow Bar.

When the **Report** button is pressed the following will automatically occur:

- Submission of sample and aberration information to the CytoSure Interpret Software database.
- Saving of data as a .cgh file to the location specified in **Tools -> Options... -> Files -> Automatically Save?**.
- Generation of a report. The format of the report can be customised but the default configuration generates a .pdf file saved in the location specified in **Tools -> Options... -> Files -> Automatically Save?**. The report contains details for each array processed such as array barcode and QC metrics. Aberration details are also generated including size, location and genes contained within each aberration.

Clicking on the **Finish** button () at any point during workflow will give the user the option to remove the dataset and, if required, save any changes. This enables the user to continue from the same position within the workflow mode. To resume a data set via workflow mode, select **Load** from the accelerate workflow bar, then select **Batch Processed Files**.

For details on customising the report type or content generated through the menu buttons, see the Other options section.

Alternative options

It is also possible to generate a report using the **Export Aberration Report...** option in the **Tools** drop-down menu (Figure 23). This will bring up the **Report Type Selection** window, as seen in Figure 24, from which the user can choose **PDF, HTML** or **Tab-delimited Text** report format. Upon choosing the report format, the **File Explorer** will open to select the save location and filename.

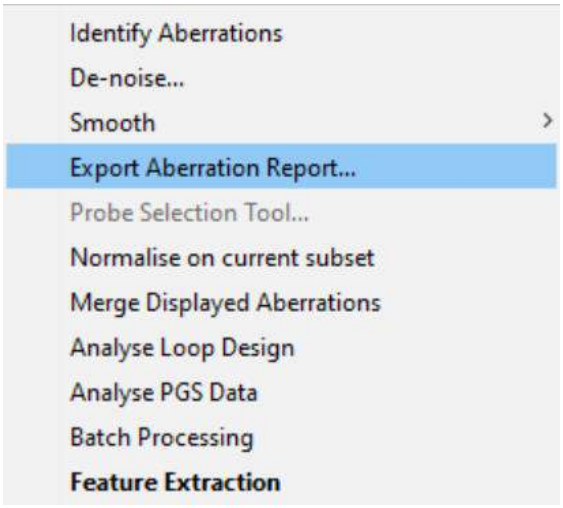


Figure 23: Export Aberration Report option in Tools drop-down menu.

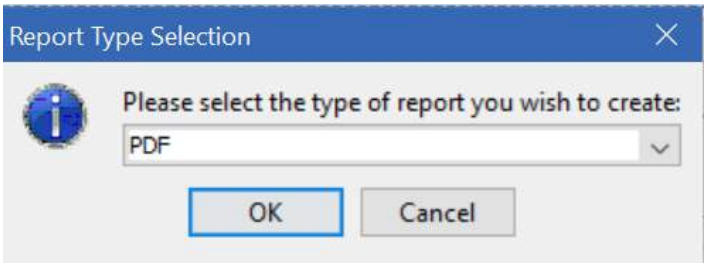


Figure 24: Report Type Selection window

The **Report Type Selection** window can also be accessed in a few other ways:

1. Clicking the **Export** button at the bottom of the **CNV Regions** table in the **Genomic View** tab (Figure 25).
2. Clicking the **Aberration Report** button on the **CNV/SNPs** tab (Figure 26).
3. On the **Database Management** tab, select a sample from the sample menu and then select the **Regenerate Report** button ( ).

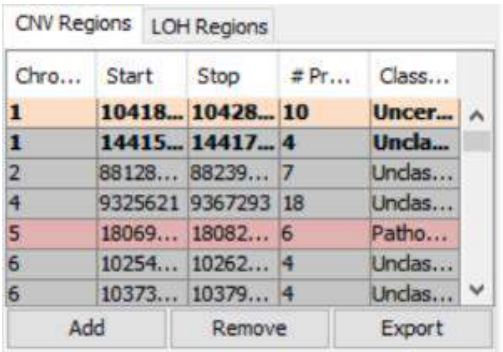


Figure 25: CNV Regions table from the Genomic View tab.

Figure 26: Aberration Report button on the CNV/SNPs tab.

Snapshots

In addition to generating reports, data can be exported from CytoSure Interpret Software using snapshot of the graphs and images in the software. There are two ways to do this:

1. To save a png image of the **Overview** and **Chromosome Sections** of the **Genomic View** tab, select **File -> Export -> Current Selection Image** and then choose a save name and location. An example of the resulting image can be seen in Figure 27.
2. To save an image of a different section of the display, right-click on an area which is not a table, graph or any other infographic. When a suitable area is clicked, a menu bar will appear (Figure 28) and some of the display will be greyed out. To save or copy the highlighted section, select **Save Screenshot...** or **Copy to Clipboard**. To select a different portion of the display, use the **Up** and **Down** options.

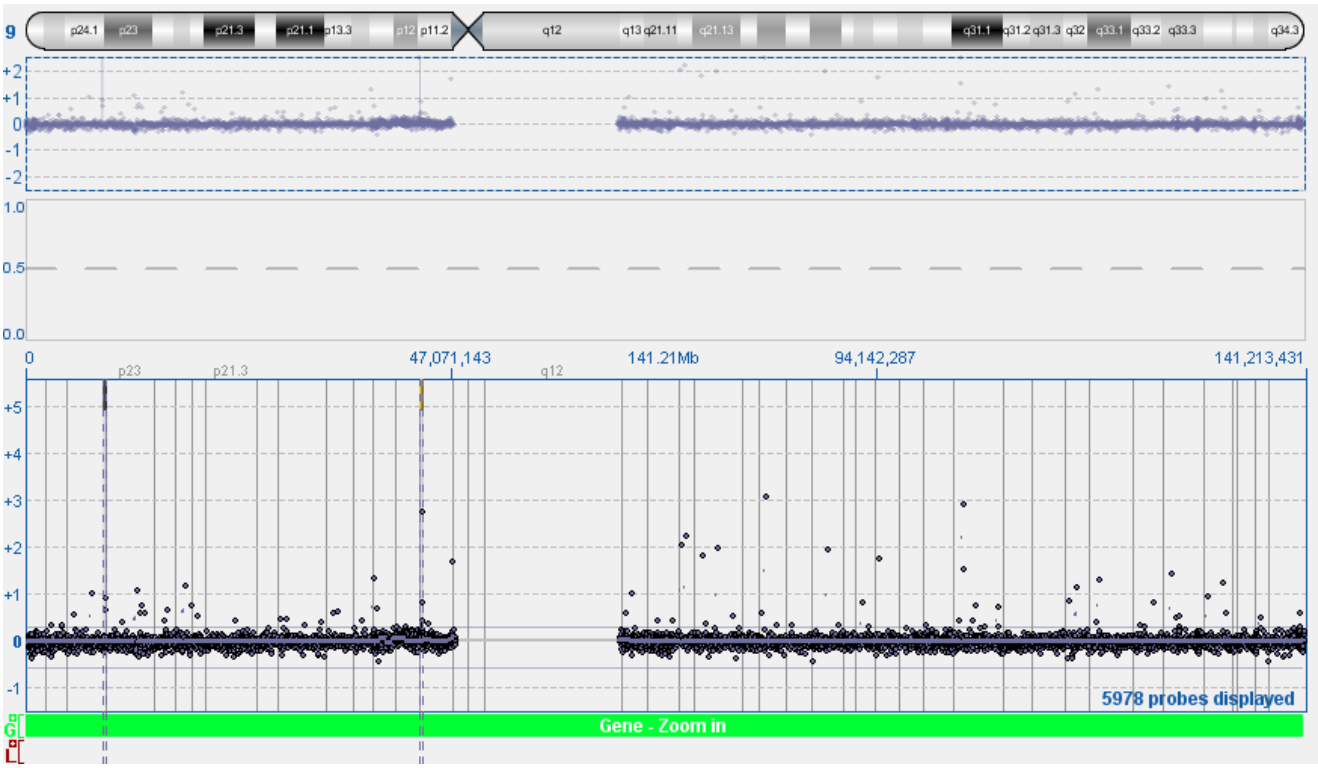


Figure 27: Snapshot image of Overview and Chromosome Sections of the Genomic View tab.

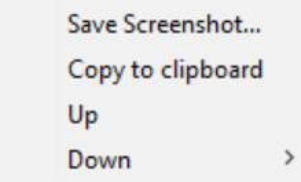


Figure 28: Screenshot Menu.

The HTML and PDF reports include CNV snapshots which are generated with the initial report or database submission and subsequently stored in the database. It is possible to regenerate these snapshots to enable, for example, the inclusion of additional annotation tracks. In order to generate these snapshots for a particular CNV or group of CNVs, select the appropriate row(s) of the table in either the **Aberrations** section of the **Genomic View** tab or the **Detected CNVs/SNPs** section of the **CNVs/SNPs** tab (ensure that **Table View** is selected), then right-click on the selection and click **(Re)Generate Snapshot(s)**. Subsequent report generation or database submission will include the latest snapshots.

Legal information

CytoSure microarrays

This product is provided under an agreement between Agilent Technologies, Inc. and OGT. The manufacture, use, sale or import of this product may be subject to one or more of U.S. patents, pending applications, and corresponding international equivalents, owned by Agilent Technologies, Inc. The purchaser has the non-transferable right to use and consume the product for RESEARCH USE ONLY AND NOT for DIAGNOSTICS PROCEDURES. It is not intended for use, and should not be used, for the diagnosis, prevention, monitoring, treatment or alleviation of any disease or condition, or for the investigation of any physiological process, in any identifiable human, or for any other medical purpose.

Trademarks

Trademarks: OGT™, CytoSure™ (Oxford Gene Technology); Bench™ (Cartagenia); NanoDrop™ (Thermo Fisher Scientific); Excel® (Microsoft Corp.).

Contact us

Oxford Gene Technology,

Begbroke Hill,

Woodstock Road,

Begbroke,

Oxfordshire

OX5 1PF, UK

T: +44(0)1865 856826; (US: 914-467-5285)
E: contact@ogt.com; W: www.ogt.com

Technical support: E: support@ogt.com
T: +44(0)1865 856826

Item number: 990241