Bionano VIA™ Software User Guide

DOCUMENT NUMBER:
CG-00043

DOCUMENT REVISION:
A

Effective Date:
07/20/2023

bionano

Table of Contents

Legal Notice

For Research Use Only. Not for use in diagnostic procedures.

This material is protected by United States Copyright Law and International Treaties. Unauthorized use of this material is prohibited. No part of the publication may be copied, reproduced, distributed, translated, reverse-engineered or transmitted in any form or by any media, or by any means, whether now known or unknown, without the express prior permission in writing from Bionano Genomics, Inc. Copying, under the law, includes translating into another language or format. The technical data contained herein is intended for ultimate destinations permitted by U.S. law. Diversion contrary to U. S. law prohibited. This publication represents the latest information available at the time of release. Due to continuous efforts to improve the product, technical changes may occur that are not reflected in this document. Bionano Genomics, Inc. reserves the right to make changes to specifications and other information contained in this publication at any time and without prior notice. Please contact Bionano Genomics, Inc. Customer Support for the latest information.

BIONANO GENOMICS, INC. DISCLAIMS ALL WARRANTIES WITH RESPECT TO THIS DOCUMENT, EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO THOSE OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. TO THE FULLEST EXTENT ALLOWED BY LAW, IN NO EVENT SHALL BIONANO GENOMICS, INC. BE LIABLE, WHETHER IN CONTRACT, TORT, WARRANTY, OR UNDER ANY STATUTE OR ON ANY OTHER BASIS FOR SPECIAL, INCIDENTAL, INDIRECT, PUNITIVE, MULTIPLE OR CONSEQUENTIAL DAMAGES IN CONNECTION WITH OR ARISING FROM THIS DOCUMENT, INCLUDING BUT NOT LIMITED TO THE USE THEREOF, WHETHER OR NOT FORESEEABLE AND WHETHER OR NOT BIONANO GENOMICS, INC. IS ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Patents

Products of Bionano Genomics® may be covered by one or more U.S. or foreign patents.

Trademarks

The Bionano logo and names of Bionano products or services are registered trademarks or trademarks owned by Bionano Genomics, Inc. (“Bionano”) in the United States and certain other countries.

Bionano™, Bionano Genomics®, Saphyr®, Saphyr Chip®, Bionano Access™, VIA™ software, and Bionano EnFocus™ are trademarks of Bionano Genomics, Inc. All other trademarks are the sole property of their respective owners.

No license to use any trademarks of Bionano is given or implied. Users are not permitted to use these trademarks without the prior written consent of Bionano. The use of these trademarks or any other materials, except as permitted herein, is expressly prohibited and may be in violation of federal or other applicable laws.

© Copyright 2023 Bionano Genomics, Inc. All rights reserved.

Revision History

REVISION	NOTES
A	Initial release.

Introduction

Variant Intelligence Applications™ (VIA) software is a complete and integrated solution for the visualization, interpretation and reporting of genomic variants from multiple technology types. By supporting multiple genome-wide data modalities, VIA™ software provides the most comprehensive view of genomic variants of any interpretation, annotation, and reporting software tool available. As a platform-agnostic tertiary analysis solution, VIA stores and manages distinct types of genomic data from various platforms (see **Table 1**) enabling the extraction of meaningful insights from a combined analysis. The software includes algorithms to detect copy number variants (CNV) from major microarray vendors, optical genome mapping (OGM), and next generation sequencing (NGS) methodologies as well as Absence of Heterozygosity (AOH), from data types that assess B-allele frequency. VIA software also provides intelligent interpretation assistance to analyze CNVs, Loss of Heterozygosity (LOH) and Structural Variants (SV) from OGM data. As a centralized analysis solution spanning technologies and application areas, VIA provides an efficient environment to keep pace with advancements in technology while retaining access to historical platform data. By being adaptive to whichever technology is used to generate CNV, LOH, or SV genomic variants, VIA software provides rich annotations for the co-analysis of sequence variants from NGS to provide a complete picture of genomic variation and reveal more answers for disease association.

Table 1. Common platforms supported in the software.

PLATFORM	EXAMPLE ASSAYS	ASSOCIATED FILE TYPES
BIONANO	OGM	ogm.bam
		ogm.vcf
THERMO FISHER/ AFFYMETRIX	Affymetrix arrays output .cel file format	.cel
	CytoScan 750K	.cychp
	CytoScan HD	.cyhd.cychp
	CytoScan XON	.xnchp
	OncoScan	.oschp
	CytoScan HT-CMA, SNP6	.cel
ILLUMINA	CytoSNP12, CytoSNP850K, Infinium Omni, GSA, GSA-Cyto, GDA, GDA-Cyto	.txt (Final Report files)
		.gtc
	Infinium HumanMethylation450,MethylationEPIC	.idat
AGILENT	SurePrint G3 CGH + SNP Bundle, 4x180K GenetiSure Cyto 4 x 180K CGH+SNP	.txt
NGS	WGS, WES, Panels	CNV = .bam
		Seq Var = .vcf, .vcf.gz .json.gz (generated by Nirvana)
CUSTOM	Custom CNV with probe or segment values	.txt (tab delimited)
	Custom Seq Var with annotations	.vcf

General overview of VIA concepts

VIA software has been designed to analyze data from a wide variety of platforms including microarray, NGS and OGM. To process data of different types, a user must have both the appropriate Sample Class enabled by their specific license and the underlying Sample Types configured in VIA. The following framework is described below (see **Figure 1**).

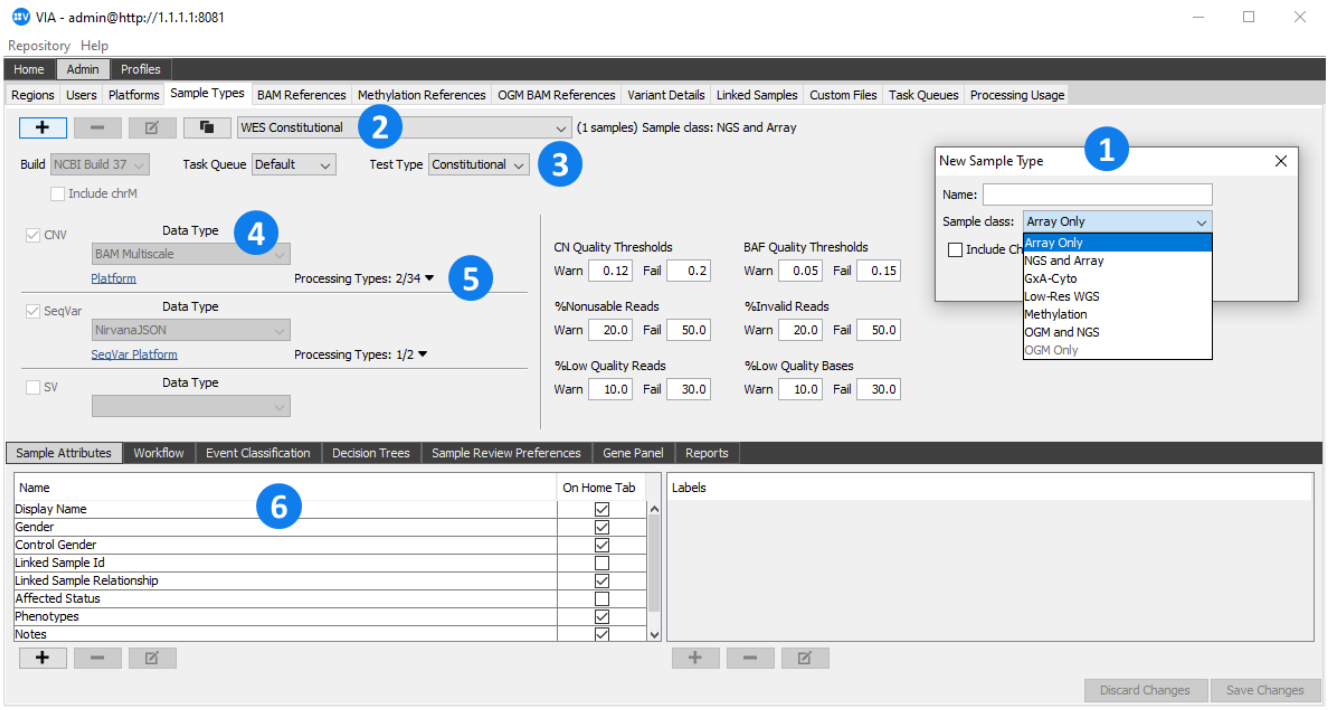


Figure 1. Numbered VIA framework.

1. **Sample Type:** The basic category for each sample loaded in VIA software. Sample type is configured by a VIA administrator under which a sample class and test type must be specified. Sample type can be used to filter samples in the database, in a **Similar Previous Cases** query, or to build aggregates or profiles. The sample type is defined by:
   - Sample Class
   - Genome Build
   - Test Type
   - Data Type
   - Sample Attributes and Analysis Preferences
2. **Sample Class:** Defines the type of input data imported from a platform or technology. Sample classes currently available in VIA include Array only, NGS and Array, GxA Cyto, Low-Res WGS, Methylation and OGM and NGS. The sample classes supported by each instance of VIA are dependent upon the user's VIA license.
3. **Test Type:** Workflow-driven parameter saved as part of the configuration of a sample type and set as either Constitutional or Oncology. Each value is associated with specific analysis features such as Parental

analysis, ACMG 2019 recommended CNV scoring, Constitutional or Oncology Knowledgebase templates, and genomic scar scoring.

4. **Modality/Data Type:** Three different modality events - CNV and/or Sequence Variants (SeqVar) and/or SV – can be used to analyze a given sample type, the respective settings specified under each event.
5. **Processing Type:** A set of parameters that converts raw data input into CNV, SeqVar or SV information. Processing types are accessible under **Platforms**. Multiple processing types can be associated with each sample type.
6. **Sample Attributes, Workflow, and Event Classification:** Examples of related components that can be defined and customized for a given sample type.

Principles of CNV and AOH Detection

Copy Number Variants (CNVs) and Absence of Heterozygosity (AOH) events can be calculated from array, NGS and OGM data. The methods used in these calculations are based on a Hidden Markov Model for Region Segmentation, a statistical probability model based on unobserved *hidden* truth sets. For CNV analysis, the truths are copy number (CN) state estimations. For each data point, predicting which state is most likely the best match by comparing the observation to the truth set/training algorithm is the objective. Numbers represent theoretical CN states. Lines represent probability calculations that a data point will change CN state.

Segmentation Algorithms

Raw data is converted into CN calls or allelic events. The presence of a signal at a specific genomic location from two different sources is measured. In the case of 2-color array comparative genome hybridization, there are two samples, one the experimental (sometimes referred to as test) sample and the second, a control (sometimes referred to as normal or reference) sample. When using single-channel arrays, such as high-density single nucleotide polymorphism (SNP) arrays, the control is a measure of signal from a large pool of samples. Regardless, it is important to note that the first step in making the copy number call is to arrive at measurements that represent the ratio of signals from the experiment as compared to control sample at multiple locations along the genome of interest. This is called the preprocessing step.

PREPROCESSING OPERATIONS

Certain data types need to undergo preprocessing steps before CN estimation can be performed. Specific preprocessing steps are unique to each technology. For data types such as ImaGene, VIA uses intensity values to arrive at log2 ratios and therefore performs preprocessing steps such as removal of flagged spots, background correction, normalization, and combining replicates. To assess reliability of the image-quantified data, many software platforms use flags to indicate areas which may be of suspicious quality and therefore removed from consideration. If a Data Type requires preprocessing operations, each has its own set of steps.

COPY NUMBER ESTIMATION

Once log2 ratios are obtained and preprocessing is complete, VIA software will arrange the ratios according to their position along the chromosome. Each probe is represented as a small gray dot along the length of a chromosome in the genome and chromosome plots; the user-specified calling thresholds seen as blue and red horizontal lines, call certain regions as a **Gain**, **Loss**, **Amplification**, or **Homozygous Loss**. VIA offers multiple

segmentation algorithms: a circular binary segmentation (CBS)-based, SNP Rank, and two hidden Markov model (HMM)-based algorithms, SNP-FASST2 and the latest SNP-FASST3. See the *VIA Theory of Operations* for detailed information on the principles of each algorithm.

Best Practices for Changing Algorithm Settings

The best way to adjust the algorithm settings is to reprocess an existing sample with modified parameters to determine the optimal setting for the application type. Since the same data file is being used, this does not tally against a user's sample count. Simply duplicate the current processing type and amend the settings; then activate the new processing type for the sample type (see **Figure 2**). Duplicate the samples of interest and process with the new processing type to compare to the original settings.

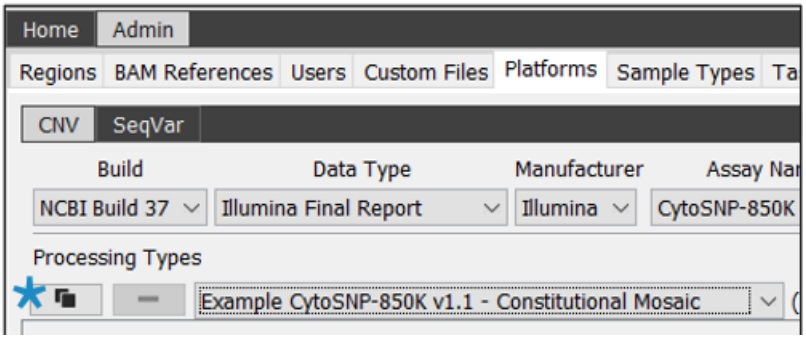


Figure 2. Processing types with amended settings.

Contrast Quality Control (QC) and segmentation performance between processing settings using the multi-sample view, as shown in **Figure 3**.

Figure 3. Segmentation performance.

For each segmentation algorithm, a significance threshold needs to be set in the Analysis panel so the sensitivity can be adjusted. The smaller the number, the less sensitive the algorithm is in creating a new segment. So, if some known aberrations are not being called because they are too small, this value should be increased. This setting is inversely proportional to the number of probes: the larger the number of probes, the smaller the value used for this setting, ensuring valid results. Many probes at a setting of 1E-6 or lower have been processed.

Home Page: Main Dashboard

The main VIA software interface, displayed in **Figure 4**, allows users to make queries to retrieve samples and load new samples. The interface is set up like a browser with labeled tabs such as **Home**, **Platforms**, and **Sample Types**.

NOTE: Upon logon, if a user license is to expire within thirty days, a message in red will be displayed under the query box. The message will disappear after the first query.

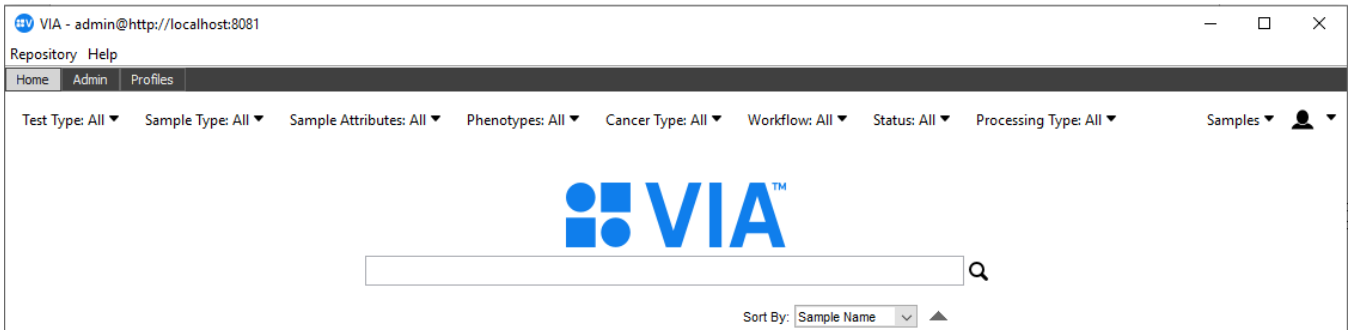


Figure 4. Main dashboard interface.

Sample Retrieval

A single multi-faceted field is used to query samples and there are multiple ways to search. To list all samples, click on the magnifying glass or hit **Enter**. A list of all samples in the database along with information on each sample will be returned, as seen in **Figure 5**.

- Search by Keywords, Attributes, Event type.
- Leave the search field empty and click the **magnifying glass** icon to list all samples.
- Filter search results by sample type, processing type.

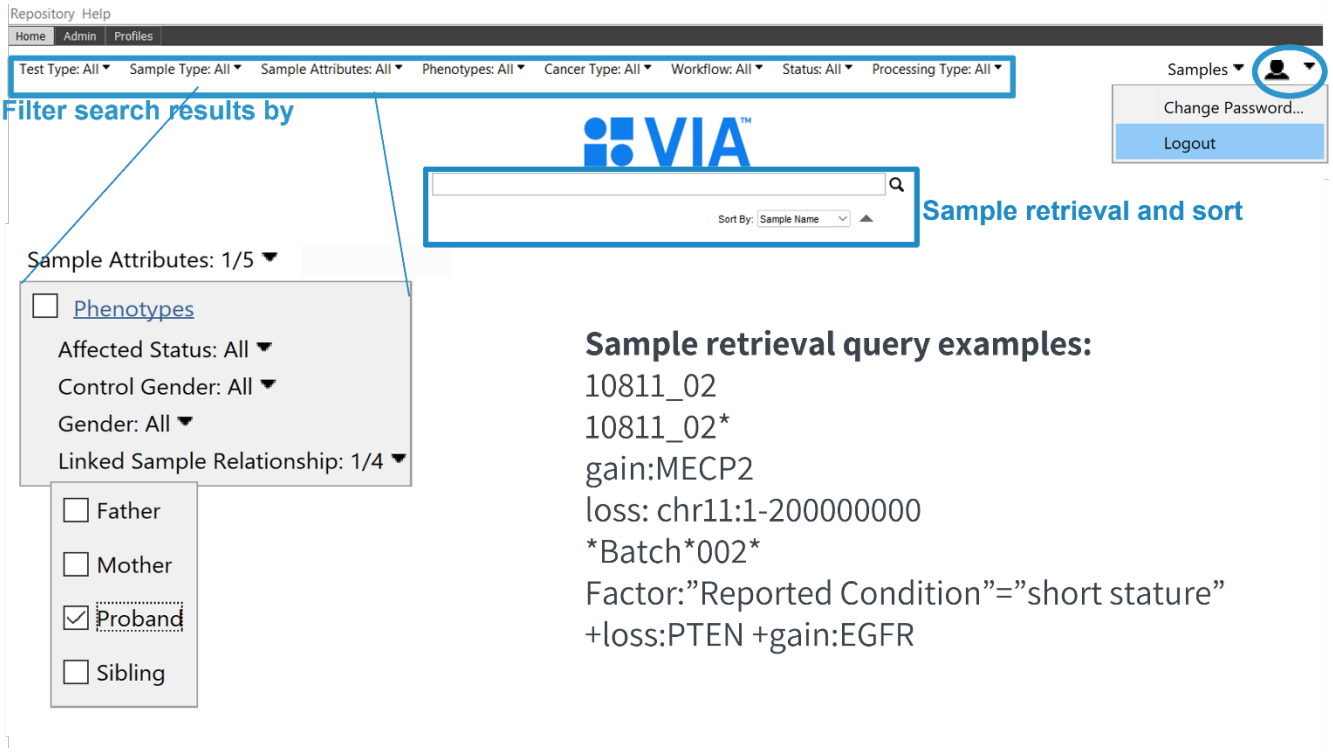


Figure 5. Sample retrieval interface.

Twenty samples are displayed on each page in alphabetical order by sample name and with various metrics. To view additional samples, click on the blue number links on the bottom right of the window to go to the additional results pages.

Each sample shows the Status (e.g., Processed, Pending) as well as the quality of the sample and the percentage of probes that were discarded from analysis. The sample type, processing type and decision tree (if applied) are shown. At the bottom of each sample displayed, the number of events classified into each category defined for the sample type (e.g., **Benign**, **Likely Benign**, **VUS**, **Likely Pathogenic**, **Pathogenic**) will be listed. Clicking on the blue sample name will open that sample in a new tab.

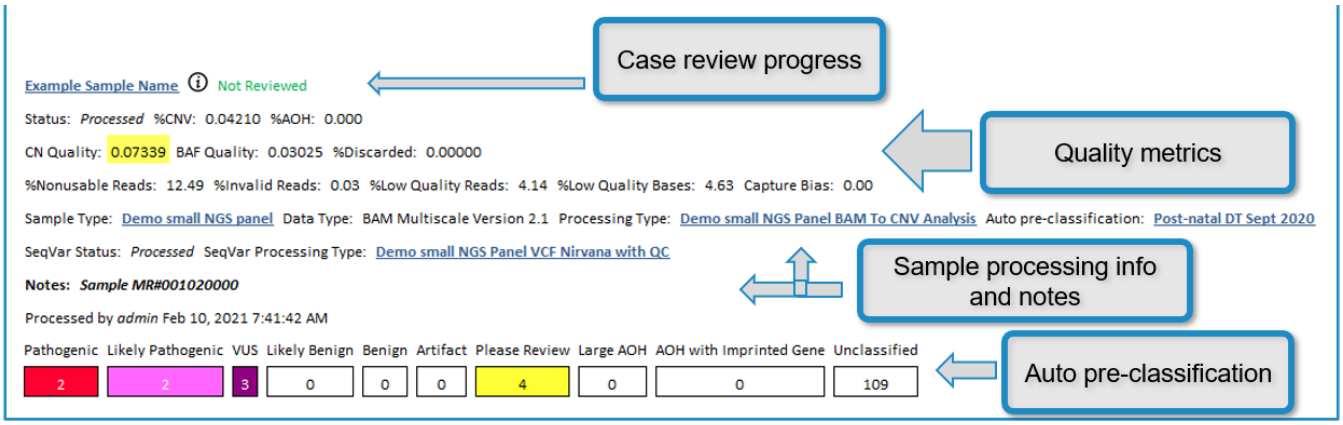


Figure 6. Sample Info.

If a sample is currently being edited by another user, the **Edit Sample** icon will be displayed next to the sample in the search results.

Moving the mouse over the icon will display the user currently editing the sample. If one is reviewing a sample and tries to edit a sample that is already being edited by someone else, an alert (in a pop-up window) will be displayed stating that another user is currently editing the sample.

If a sample is under a sample type with the test type specified as oncology, then the icon will appear to the left of the sample name. Clicking the icon directly launches the **Profile** window for the **Oncology Profiles** feature.

Quality Metrics

Sample metrics are displayed for each sample in the **Home** page query results.

These scores may be highlighted in yellow or red, as shown in **Figure 7**, on the **Home** page results indicating they exceeded thresholds set by the Admin for this sample type. This CN Quality score is also used during processing to remove probes from the probe median calculation. The percentage of outliers to remove under the **Robust Variance Sample QC** (see next section) parameter in **Settings** specifies what percent of probes to remove.

>= Warn threshold	CN Quality: 0.11545

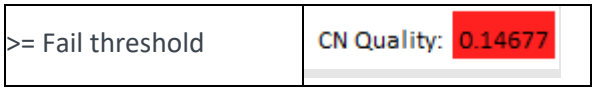


Figure 7. Highlighted scores.

- **%CNV:** % of the genome (excluding sex chromosomes) that has CN changes. Calculated as total length of autosomal regions with CN changes divided by the total length of the autosomes.
- **%AOH:** % of genome (excluding sex chromosomes) with AOH calls. Calculated as total length of AOH regions over autosomes (excluding those overlapping CN changes) divided by the total length of the autosomes. The min LOH length threshold for each sample type is set by the admin under **Processing Type**. Based on the minimum LOH setting of 2,500KB, and on studies showing % homozygosity is correlated with some degree of consanguinity, a 6% AOH threshold is used as a warning indicator. If the %AOH is over 6%, the value will be highlighted in yellow.
- **BAF Quality:** % of SNP probes with BAF values in the allelic imbalance region of the BAF track. This quality score indicates how well SNP probes are behaving. It reports the percentage of SNP probes with BAF values between the heterozygous imbalance threshold and homozygous value threshold as defined in the **Processing** settings.
- **CN Quality:** A score representing the probe-to-probe variance measuring on average how much successive probes differ from each other, shown in **Figure 8**. Please note that these scores are only used to compare relative CN Quality scores between samples within a platform. It can help indicate reliability of the data; a higher CN Quality score indicates less reliable data while a lower CN score indicates more reliable data. A high score may indicate low quality DNA, debris on the array slide, or other issues during wet lab prep.

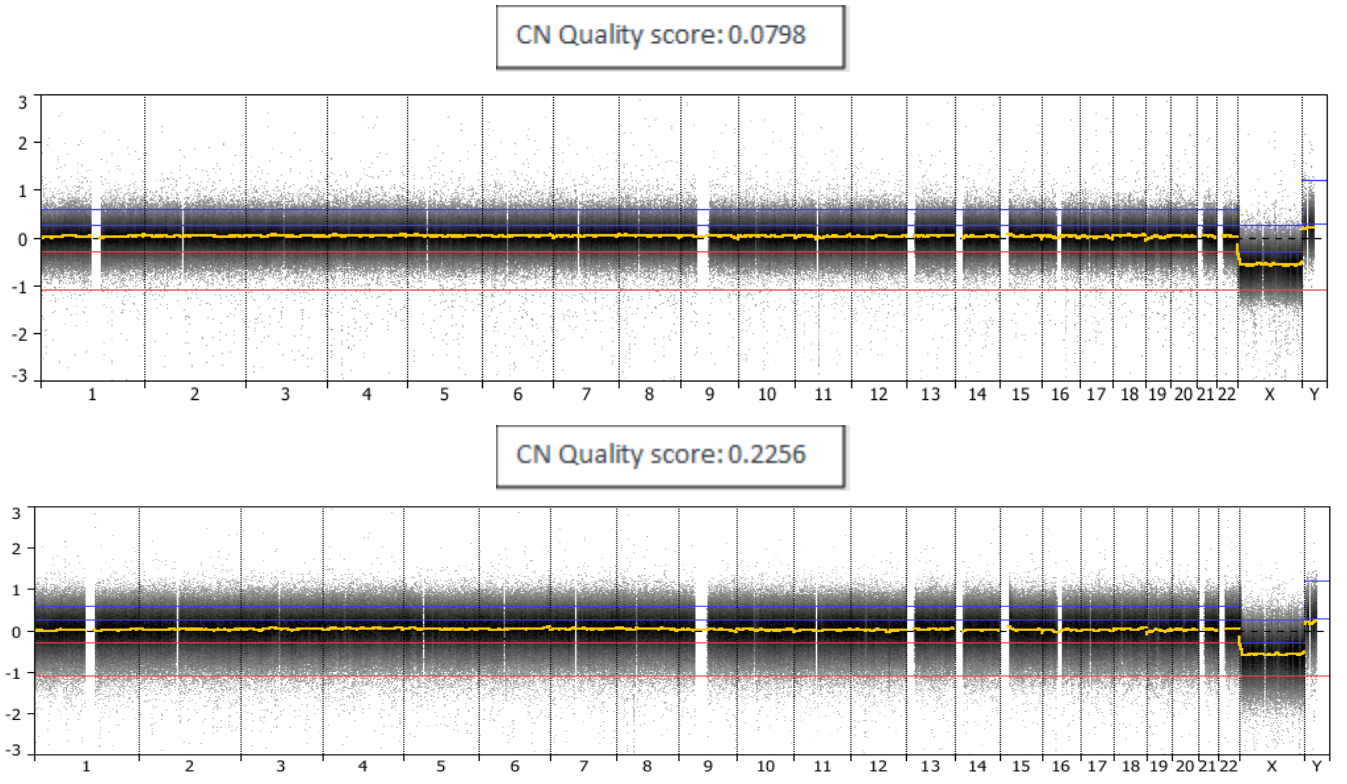


Figure 8. Example probe plots with two different quality scores. Probes for the sample with the smaller quality score (0.0798) are packed more tightly around the 0 than in the sample with the larger CN Quality score.

ROBUST VARIANCE SAMPLE QC CALCULATION – QUALITY SCORE

This **Quality** score is a calculated value representing the probe-to-probe variance between adjacent probe log ratios after excluding the outliers. This single parameter is used to remove from calculation of the variance, the extreme outliers that one would expect due to copy number breakpoints. It is meant to measure how much successive probes differ from each other on average. The score is displayed as **Quality** in the **Home** page results. The score is computed by first ordering by magnitude the difference between adjacent probes and then removing a percentage of the probes that fall at the top and bottom of the list. For example, if probes ordered along the genome have values [1, 2, 1.5, 2.1] then the differences would be [1, -.5, .6]. Then a percentage of the probes would be removed (from calculation of the mean variance) from the top and bottom of the variance spectrum. If the value specified is 0.2 (for 0.2%), then half of this percentage of probes (0.1%) are removed from the top of the list and the other half, from the bottom. The percentage to remove is set by the Admin in the **Processing** settings (Percent outliers to remove).

The default value for most processing types is 0.2% but can be changed individually for each processing type for which QC calculation is available. A good starting point for the setting would be a calculation such as: 2*(expected number of CNVs)/(number of probes). For example, if one expects a maximum of 500 CNVs and an array with roughly a million probes, the value can be set to 1000/1M or 0.1%.

Please note that these scores are only used to compare relative QC scores between samples and the outliers are not removed from processing. The lower the QC score, the better the quality (low variance). One can visually see the difference between samples with high and low QC scores as in **Figure 9** and **Figure 10** below. The sample with a higher QC score (**Figure 10**) displays a lot more noise as compared to the sample with the lower QC score.

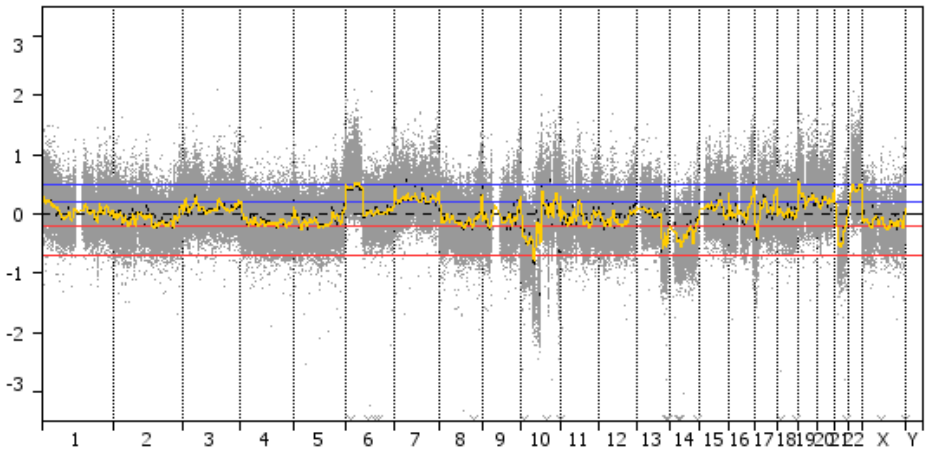


Figure 9. Sample with lower QC score: 0.09.

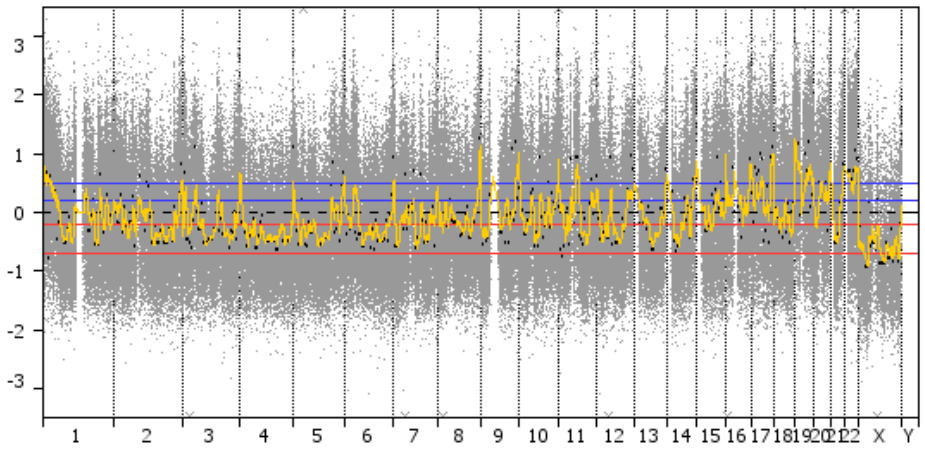


Figure 10. Sample with higher QC score: 0.67.

Having the QC value gives users a better idea of how much confidence to have in the sample's aberration calls. If a sample's scores are too high, one can re-run the assay. A quality score of 0.15 – 0.2 is generally considered the cut-off range for good samples for most arrays.

BAM Quality Control Metrics

QC metrics for BAM files associated with CNV or SeqVar are displayed on the **Home** page results. If read quality data is missing from the BAM files, these values will be blank. All values are reported as a percentage of reads.

%Nonusable Reads: 5.71 %Invalid Reads: 0.20 %Low Quality Reads: 0.00 %Low Quality Bases: 2.29

These scores may be highlighted in yellow or red on the Home page results indicating they exceeded thresholds set by the Admin when creating a sample type.

>= Warn threshold	%Low Quality Reads: 14.77
>= Fail threshold	%Low Quality Reads: 38.97

Metric Calculation and Description

%Nonusable reads: (Total Reads - Usable Reads)/Total Reads * 100. Nonusable reads are all read map/align records that were discarded and not used in the depth counting. These include reads flagged in the BAM file as Unmapped BAM reads, Duplicate BAM reads, Secondary BAM alignments, Clipped BAM alignments, and reads classified as Invalid reads (see below).

%Invalid Reads: Invalid Reads/Total Reads * 100. Invalid reads include reads with undefined reference sequences (e.g., alternate haplotype seqs) and reads where the start/end coordinates do not match. In some BAM files, the alignment flags will indicate that a read should have valid mapping but then the chromosome/reference-sequence or start and end positions will be undefined in the alignment record. In other cases, the set of reference sequences used by the aligner to map reads includes extra contigs or decoys or

alternate haplotype sequences that are not included in the VIA standard reference genome. Reads mapped to these extra sequences will also be discarded in this category as not having valid BAM coordinates.

%Low Quality Reads: Number of Reads with MAPQ<30/Total Reads * 100. MapQ<30 looks at the number of read mapped/aligned records for which the aligner gave a score below Phred 30 (more than 1/1000 chance of mapping position being wrong).

%Low Quality Bases: Number of Sequence Bases with Quality <30/Total Sequence Bases * 100. BaseQ<3 looks at the number of raw read base calls that the sequencer gave a quality score below Phred 30 (more than 1/1000 chance of being wrong).

Find… (Searching Results on Page)

The search field at the top right of the window, seen in **Figure 11**, (next to the **Samples** dropdown) is used to find specific text within the displayed page (like the **Find…** tool in a browser window). For example, to find samples that were uploaded on July 7, enter Jul 07 in the **Search** box, and click on **Magnifying Glass** or hit **Enter**. If found, the search term will be highlighted in yellow on the page. Hitting the **Enter** key successively will highlight the next occurrence of the search term, shown in **Figure 12**.

Figure 11. The Search field.

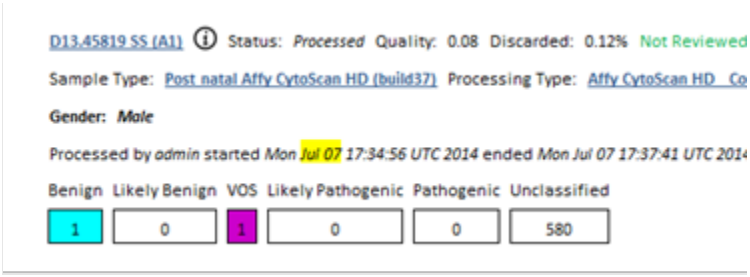


Figure 12. Example results.

NOTE: Wildcards cannot be used in this field. This only searches the displayed page, not all results from a query. The field is only displayed when samples are returned and displayed on the page from a query. The field is hidden if no samples result from a query.

Searching/Querying the Sample Database

To list only specific samples, enter the query into the **Search** field in the middle of the window. Samples can be queried in multiple fields including sample name, genomic region, gene symbol, cytoband, sample attribute, processing date, user who processed sample, or a combination of these.

Search/Query Inputs

Quotation marks: When quotation marks are used in queries, ASCII quotes must be used. Do not use quotes from other formats, such as smart quotes. Often, word processers such as MS word use smart quotes and if copied over to the VIA query field, the query will fail. To prevent query failures, the recommendation is to type out the quotes in the query field.

Region coordinates: Commas cannot be used to specify bp (base pair) positions for the search and case matters. Location must be a chromosomal range; one cannot specify just the chromosome to include the entire chromosome. To search for an event over an entire chromosome, use a start position of 1 and end position equal to or longer than the chromosome length.

Numerical fields: Samples can be searched via numerical fields to retrieve samples containing field values greater than or less than a numerical value. Operators supported for numerical fields: <, <=, >, >=, =, !=.

Queries using conditional (AND/OR): The "**+**" sign before the query term indicates that the sample must meet the condition (AND). A query term without a preceding symbol indicates an OR statement. A "**–**" sign before the query term indicates that the sample must not meet the condition.

Examples:

Searching for sample with either a gain OR loss overlapping EGFR

```
loss:EGFR gain:EGFR
```

Searching for samples with a loss overlapping PTEN AND a gain overlapping EGFR

```
+loss:PTEN +gain:EGFR
```

Searching for samples that must contain a gain of EGFR and NOT a loss of FOXA1

```
+gain:EGFR -loss:FOXA1
```

QUERYING BY SAMPLE NAME

Samples can be searched by sample name by simply entering the sample name into the **Search** field. The * can be used as a wild card during a search by sample name. Multiple wildcards can be used in the same query.

Examples:

Search for all samples beginning with "Batch": **`Batch*`**

Search for samples that have "Batch" and the word "male" somewhere in the name: **`*Batch*male*`**

Search for samples with the number 80021: **`*80021*`**

Search for all samples ending with June2019: **`*June2019`**

QUERYING BY EVENT

Queries for the following events are accepted: Gain, Loss, AOH, and SeqVar. Acceptable syntax for events: gain, loss, cn gain, cn loss, aoh, SeqVar. Region coordinates must not contain commas and lower-case letters should be used for chr.

The basic syntax is [event]:[chromosome]:[start position]-[end position] for a bp location. To specify a cytoband, use [event]:[chromosome][band]

Examples:

Events overlapping a gene: **`gain:PTEN`**

Event overlapping a region: **`aoh:chr11:30507990-83574562`**

Events on a single chromosome (use 1 as start bp): **`loss:chr11:1-200000000`**

Events at a cytoband: **`loss:1q31.3`**

QUERYING BY SAMPLE ATTRIBUTE

One can query for a specific attribute (Factor) value by typing in Factor: followed by the factor name and value. The format is Factor:[attribute name]=[attribute value].

Examples:

Search for Male samples: **`Factor:Gender=Male`**

Search for sample where the Attribute and/or value is a multi-word term: **`Factor:"Reported Condition"="short stature"`**

Searching for all samples belonging to linked samples: **`Factor:"Linked Sample Id"=Adams`**

QUERYING BY QUALITY METRICS

To query for samples meeting a specified quality metric type in dna_attribute:[quality metric] and the threshold.

Examples:

Searching for samples with CN quality score **`dna_attribute:quality<="0.10"`**

Searching for samples with BAF quality score **`dna_attribute:snp_quality>"0.01"`**

Searching for samples with %CNV **`dna_attribute:percent_cnv>="10"`**

Searching for samples with %AOH **`dna_attribute:percent_aoh<="12"`**

QUERYING BY CLASSIFIED EVENTS AND EVENT TYPES

Queries can be performed for classified events and event types. The query returns samples with specified event type having the specified classification. Any user-defined classification value can be specified. Event types are categorized as CN change, allelic event, sequence variant, and structural variant. Additionally, CN events with a specific number of the event can be queried.

Examples:

CN event (Gain, Loss, Amplification, Homozygous Loss) classified as “Likely Pathogenic”
`sample_term:"DNAData:cn_cls_sum:Likely Pathogenic"`

Allelic event (AOH, Allelic Imbalance) classified as
“Benign:”**`sample_term:"DNAData:snp_cls_sum:Benign"`**

Sequence Variant event (SNV, Deletion, Insertion) classified as “SV in Dominant Gene”
`sample_term:"DNAData:SeqVar:cls_sum:SV in Dominant Gene"`

CN loss with 11 events **`dna_attribute:"CN Loss"=11`**

Amplification with at least 5 events: **`dna_attribute:"Homozygous Copy Loss">=5`**

QUERYING BY BENIGN AND PATHOGENIC CLASSIFIED CN EVENTS OVERLAPPING A REGION

Samples can be searched with benign and pathogenic classified CN events overlapping a region. The query returns samples with classified copy number events overlapping the specified bp location/gene. The only classification values supported are “benign” and “pathogenic”; they are case sensitive. Both regions and gene symbols are supported. Only works for CN events.

Examples:

Pathogenic CN events overlapping a region: **`pathogenic:chr13:46573687-51607314`**

Benign CN events overlapping a gene: **`benign:PTEN`**

QUERYING BY SAMPLE PROCESSING DETAILS

Samples can be searched by details submitted when the sample was loaded and processed. The system takes in the date with respect to the date/time of the VIA server. VIA assumes dates to be dates in the format yyyy-mm-dd. Other processing information available for query are the username, estimated gender at processing, decision tree, BAM MSR, and human genome build.

Examples:

Searching for samples processed after 2023-01-31, not including the date:
`+dna_attribute:proc_ended_timestamp>"2023-01-31"`

Searching for samples processed between 2020-09-10 and 2022-01-07 and on the dates:

```
+dna_attribute:proc_ended_timestamp>="2020-09-10"
+dna_attribute:proc_ended_timestamp<="2022-01-07"
```

Searching for samples loaded on 2022-12-31: **`+dna_attribute:loading_ended_timestamp="2022-12-31"`**

Searching for samples processed by username = admin: **`dna_attribute:proc_by="admin"`**

Searching for samples processed with a specific BAM MSR file: **`dna_attribute:bam-ref="PG02W_Male"`**

Searching for samples processed with a specific human genome build file: **`dna_attribute:build="NCBI Build 37"`**

Searching for samples processed with a specific decision tree: **`dna_attribute:decision_tree_name="AML DT"`**

QUERYING BY PHENOTYPES

The VIA database can be queried by phenotypes. **NOTE**: Phenotypes must be entered as HPO IDs, not text. Use * before and after the ID to indicate that any other HPO ID can be present to include samples that may contain more than the specified phenotype. If looking for samples containing only a single phenotype, do not use * before or after.

Examples:

Searching for samples that contain the phenotype “Seizures”: **`factor:Phenotypes="*HP:0001250*"`**

Searching for samples that contain the phenotype “Seizures” or “Global Developmental Delay”: **`factor:Phenotypes="*HP:0001263*"`**

Searching for samples that contain only the phenotype “Seizures” and no other phenotypes: **`factor:Phenotypes="HP:0001250"`**

FILTERING QUERY RESULTS TO NARROW LIST

Users can also filter results by one of the categories listed on the top of the window. Numbers next to the filter names show how many parameters are selected and the total number of filter fields for that category. For example, to search only for specific sample types, mark off the checkboxes of the types desired to search under the **Sample Type** dropdown. In **Figure 13** below, two out of thirty-four values are selected as indicated by the 2/34 next to sample type. If no boxes are checked off, all samples in the database are searched. Values displayed in these dropdowns are specific for each installation and are based on what has been defined by the VIA Administrator

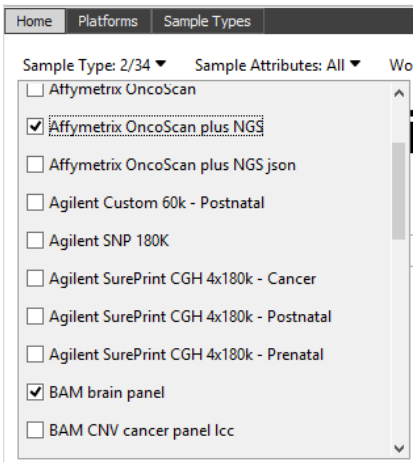


Figure 13. Sample Type, Sample Attributes, Workflow, Status, or Processing Type

VIEWING MULTIPLE SAMPLES TOGETHER

Shown in **Figure 14**, in a single window on the **Home** page, search for the samples to view together and click on the **Multi-Sample View** link.

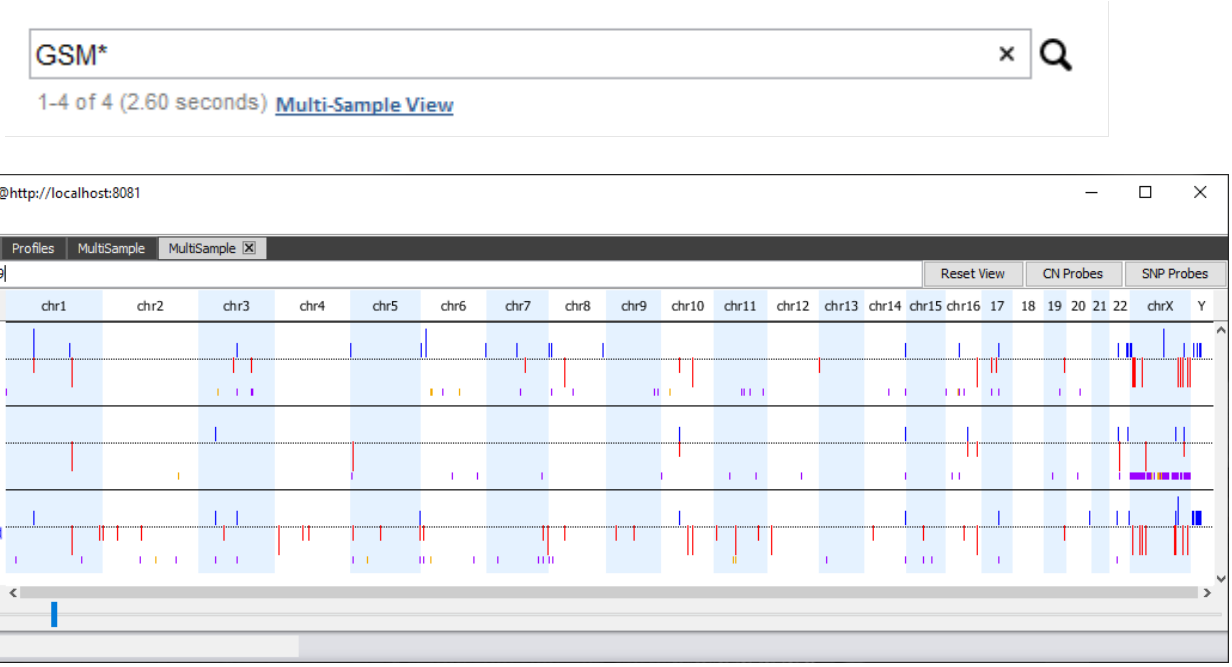


Figure 14. All samples listed on the **Home** page will be displayed in a new tab.

Clicking and dragging over the tracks or just left clicking on the labels will zoom in on the area. The top left corner displays the chromosomal range displayed; this can be edited to zoom into a specific region. The **Reset View** button on the top right will revert to the fully zoomed out view. To open one of the samples into a single sample **Review** tab, click on the blue hyperlinked sample name, displayed in **Figure 15**.

Figure 15. At sufficient zoom levels, the CN Probes and SNP Probes, toggle buttons will become active.

Clicking on the toggle buttons will display/hide the respective probes (CN or SNP), as seen in **Figure 16** and **Figure 17**. Probes will only be displayed at the sufficient zoom level; if the user zooms out of the probe level while probes are displayed, they will be replaced with the calls and the toggle buttons will become inactive.

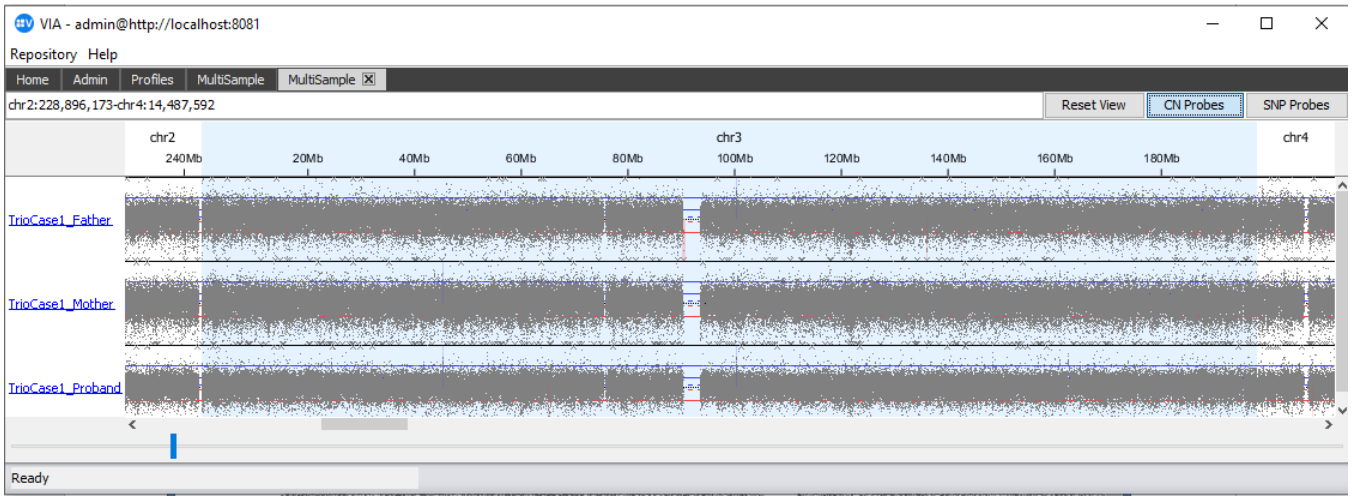


Figure 16. CN probes are displayed.

Figure 17. SNP probes are displayed.

Exporting Event Information

EXPORT EVENTS TABLE FOR A BATCH OF SAMPLES

The event table for each sample can also be exported as a separate tab delimited txt file for each sample through the **Home** page. To export sample data through this route, each of the samples to export will need to be in the locked state. The user can then query the list of samples for export and select **Samples > Export > Detailed Events Table**, as seen in **Figure 18**.

NOTE: The exported table will export only the selected events and detailed information at the time the sample was locked.

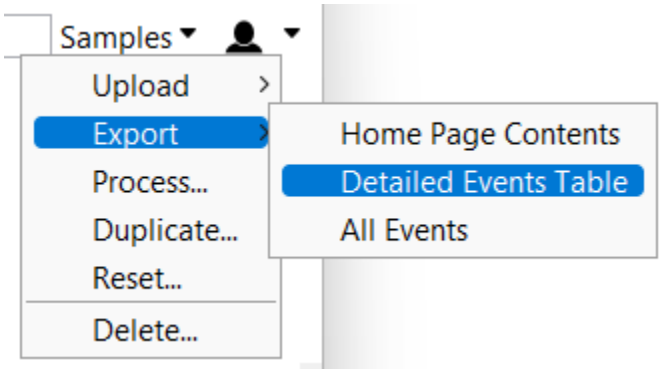


Figure 18. Events Table route.

EXPORT LIMITED EVENT INFORMATION FOR A BATCH OF SAMPLES

Limited event data for a batch of samples can be exported as a JSON file. This is done by querying for a list of samples and selecting **Samples > Export > All Events**, shown in **Figure 19**.

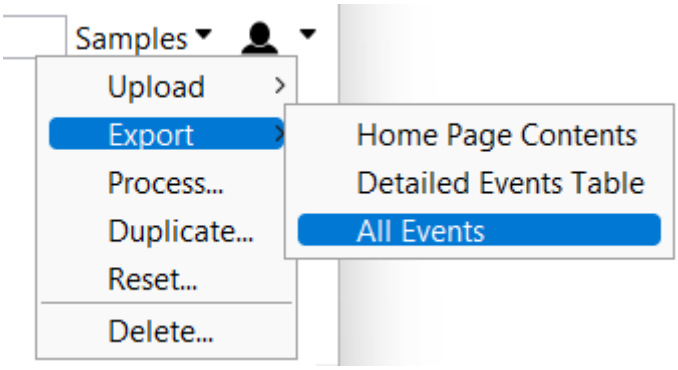


Figure 19. Limited event data route.

Caveats to this method:

- The exported data contains only the following for all samples in the query results: Sample name, Event type (e.g., CN Gain, SNV), Genomic Coordinates, Event Classification and Event Notes.
- All the events, regardless of whether the event had the select check box on, will be exported from the sample.
- The sample information (QC metrics, sample attributes, sample type designation) can be exported as a txt file as well. This is performed by querying for a set of samples and selecting **Samples > Export > Home Page Contents**.

  NOTE: This export method will not export individual event information.

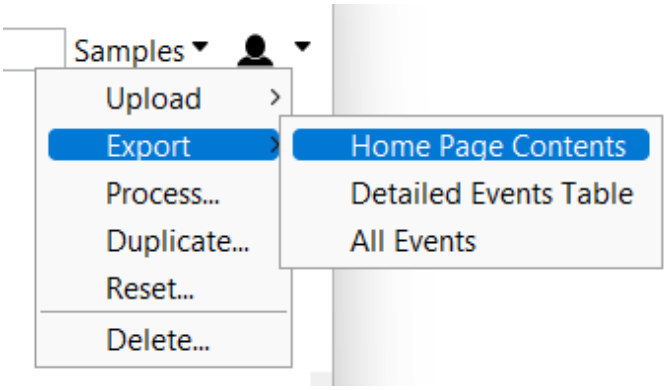


Figure 20. Export **Home** page Contents.

EXPORT EVENTS AS JSON FILES FROM HOME PAGE

All events in samples from a query can be exported into a single zipped JSON file. This file can then be read into a customer's LIMS/reporting tool. The exported data contains the following for all samples in the query results:

- The "version" refers to the JSON export schema version.
- The **Samples** field lists each sample that was returned via the **Home** page query and outputted in the JSON.
- The **Events** field lists each modality (cnvEvents, snpEvents, and SeqVarEvents) available in a particular sample.
- Sample name

- Event type (e.g., CN Gain, SNV)
- Genomic Coordinates
- Classification
- Notes

Follow the steps below to export JSON files.

1. Open a client and query for a set of samples via the **Home** page **Search** box.
2. Click on the **Samples** menu on the upper-right corner and select **Export > All Events**, (see **Figure 21**).

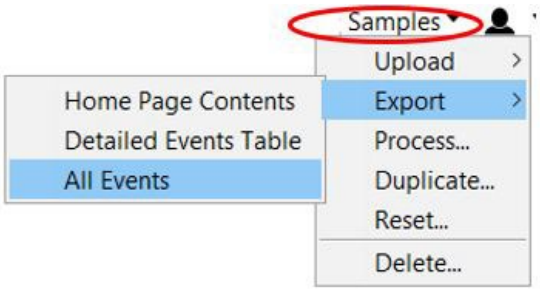


Figure 21. Dropdown menu for exporting.

3. In the resulting file chooser, select the location for the export file and provide a name.
4. Click **Open**; a progress bar window opens showing the number of samples exported and number remaining. During the export, all interactions with the client are blocked.

Once the process is complete, an alert shows that samples were exported successfully, and the data exported will be a zipped version of the JSON format described above.

Sample Review – User Functions

Clicking on the blue hyperlinked sample name opens the sample in its own tab in the window and automatically makes that sample tab active/selected. To open a sample but not make it selected, hold down the **CTRL** key while clicking on the sample name. A new tab will appear for that sample, but it will not be selected. A genome browser with annotation tracks, a results table, variant details view, and numerous tools (e.g., filtering) are available in the **Sample Review** window, as seen in **Figure 22**.

Some visual elements described below may not be displayed for each sample. These elements are dependent on the sample type of the case under review and modalities loaded with that sample.

Sample Review Overview

Across the top section of the window the genome browser with detailed variant information in an interactive and visual format is housed. This section is divided into two tabs: **Tracks** and **Variant Details**. Below this, the results

are displayed in table format. The lower section also has additional tabs for the **whole genome**, **deleted events** and **report** displays.

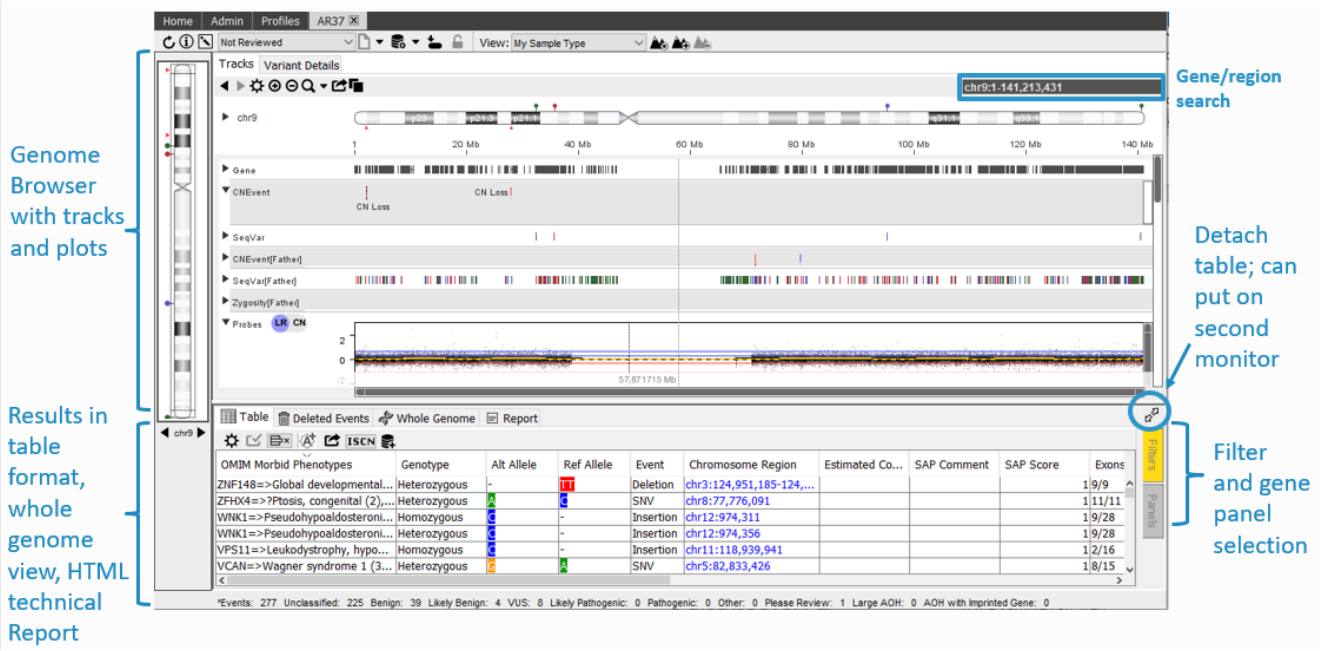


Figure 22. Sample Review window.

Sample Review Window Layout

By default, both the tabular data table and the graphical display (tracks/ideogram) are positioned one above the other in a single window, as shown in **Figure 23**. The amount of space allotted to each can be adjusted by moving the divider bar up or down.

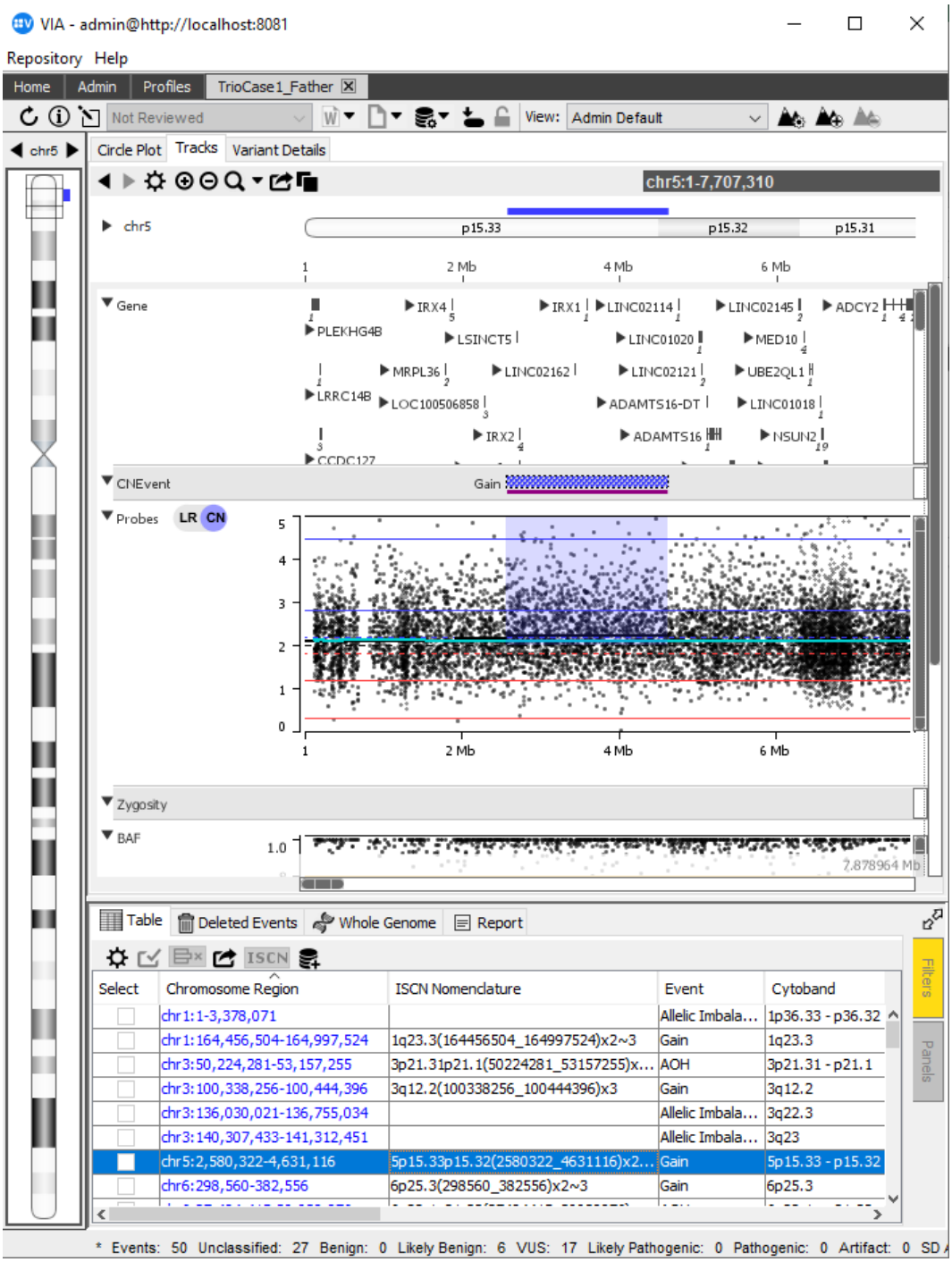


Figure 23. Tracks and ideograms.

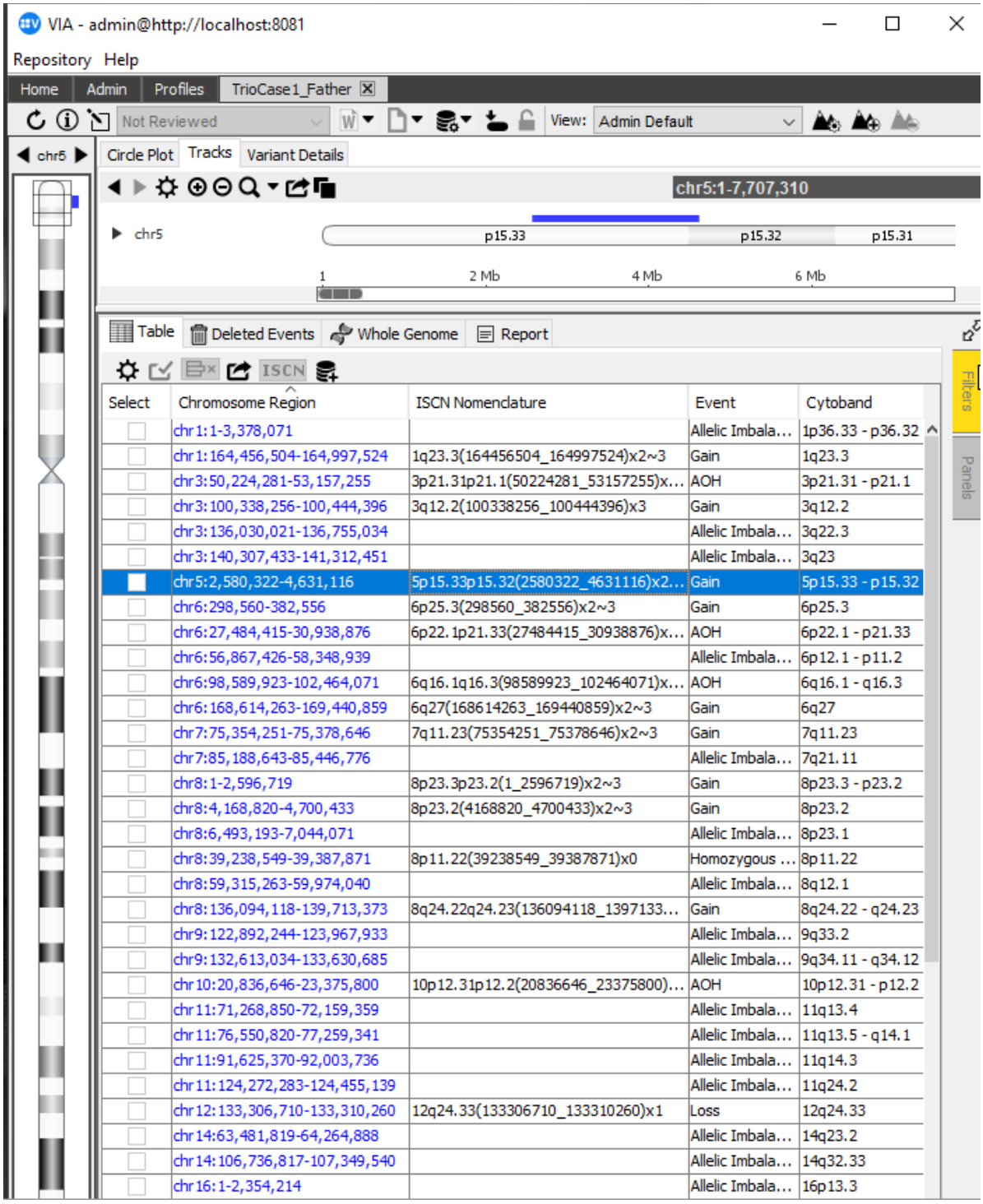


Figure 24. Repository Help

The table can be detached from the main window and placed into its own window using the black triangles on the top right of the data table.

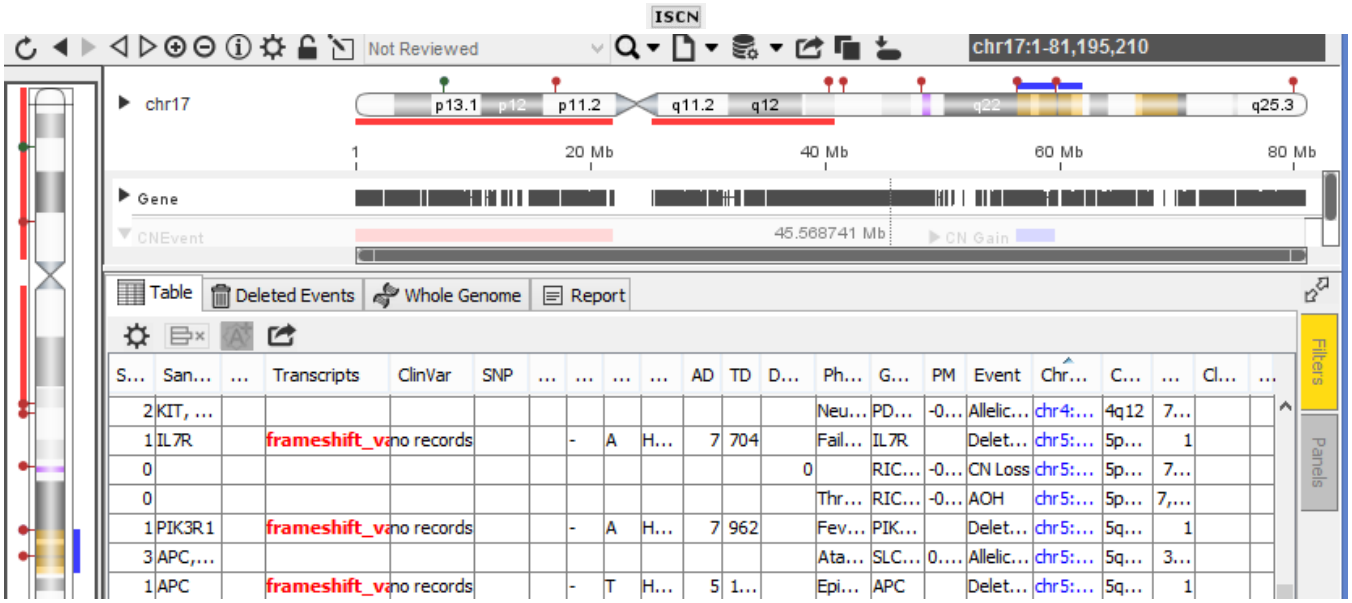


Figure 25. ISCN Screen

Once detached, the table window can be enlarged, and the two windows (ideogram and tracks table) can even be placed on separate monitors when using multiple display monitors. Clicking on the double arrows or closing the window will attach the table window back to the bottom of the main window.

CNV Events - Constitutional

For CNV events, the **Variant Details** tab is divided into three vertical panels. The left-side panel displays basic information about the variant (e.g., size, number of probes, classification, notes, audit log). A representation of Similar Previous Cases and DGV similarity is displayed as a dial. The DGV similarity is based on overlap with the curated DGV track. The curated DGV track excludes CNVs from any study including BAC results and all events less than 50bp.

The middle panel, **Gene Details**, shows the information about the gene content of the region and associated information about the genes. Additionally, for Constitutional Test Type samples, phenotype information for the relevant genes is listed. The Significance Associated with Phenotype (SAP) score will also be displayed if phenotypes were associated with the sample, otherwise, if no phenotypes were specified the displayed score will be 1 for all genes.

Gene Details Table

For the gene related information displayed in the **Gene Details** table, **Provided Regions** tracks are used to gather gene information, as shown in **Figure 26**:

- **Gene:** RefSeq Genes track; gene is hyperlinked and links out to NCBI.
- **Inheritance:** OMIM Phenotypes track; phenotype is hyperlinked and links out to OMIM.
- **OMIM Phenotypes:** OMIM Phenotypes track.

- **OMIM:** O=OMIM gene; M=Morbid gene.
- **Haplo-insufficiency:** ClinGen Haploinsufficiency tracks; B=Benign, LP=Likely Pathogenic, P=Pathogenic.
- **Triplo-sensitivity:** ClinGen Triplosensitivity tracks; B=Benign, LP=Likely Pathogenic, P=Pathogenic.
- **DDG2P:** DDG2P Confirmed and Unconfirmed tracks; CB=Confirmed Biallelic; CM=Confirmed Monoallelic, UB=Unconfirmed Biallelic, UM=Unconfirmed Monoallelic.
- **Imprinted:** Imprinted Genes tracks; P=Imprinted Genes Paternal, M=Imprinted Genes Maternal, D=Predicted Imprinted Genes, V=Provisional Imprinted Genes, ID=Imprinted Genes Isoform Dependent, R=Imprinted Genes Random.

Detailed track name and information can be found in the **Regions** tab within the **BioDiscovery Provided Regions** folder.

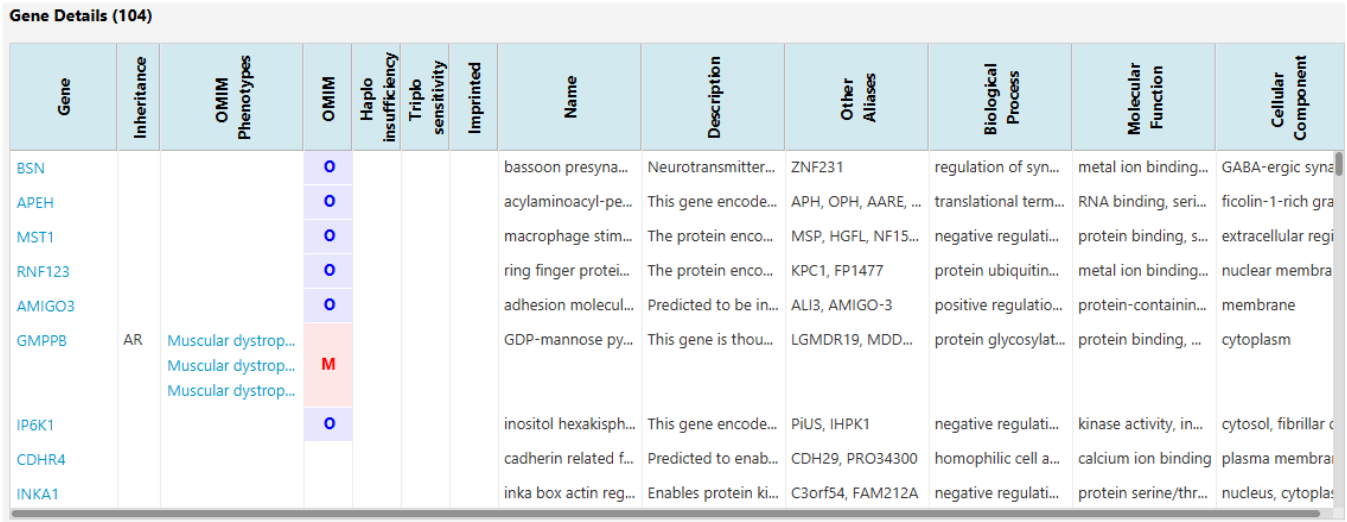


Figure 26. The gene ATAD3A with a single phenotype but with both biallelic and monoallelic inheritance with each in a different category in the **DDG2P** database, is displayed in both tracks.

The **Gene Details** columns **Haploinsufficiency**, **Triplosensitivity**, **DDG2P**, and **Imprinted** will show only a single value - the most severe/interesting clinical annotation. If a gene has different modes of inheritance with different levels of certainty for the phenotype association and mode of inheritance in the **DDG2P** database, then only one entry will be present in the table. Which one is displayed is based on the severity of the clinical annotation (more severe/interesting will be displayed). See **Figure 27**.

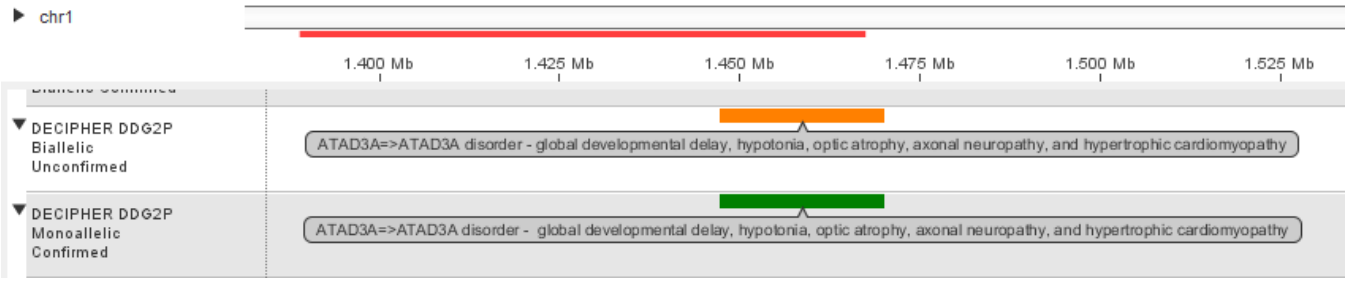


Figure 27. The **Gene Details** table only displays CM (confirmed monoallelic) in the **DDG2P** column.

The function uses a priority list to determine which annotation is more severe and therefore will be the one represented in the table when there are multiple values, as seen in **Figure 28**.

DDG2P priority list (higher priority at the top):
DECIPHER DDG2P Monoallelic Confirmed
DECIPHER DDG2P Biallelic Confirmed
DECIPHER DDG2P Monoallelic Unconfirmed
DECIPHER DDG2P Biallelic Unconfirmed

Figure 28. Priority list for Gene Details.

If phenotypes are associated with the sample, SAP score for the phenotypes and level of associated are displayed, as in **Figure 29**. However, for **Oncology Test Type** samples, no phenotype and SAP score, DDG2P, and dosage sensitivity information is displayed.

Gene	Phenotypes (SAP Score = 4.354E-13)	Significance
CFHR1	**Level 2** Seizure->Symptomatic seizures,Focal-onset seizure **Level 3** Seizure->Bilateral tonic-clonic seizure Cognitive impairment->Intellectual disability	4.354E-13
CFHR3	**Level 2** Seizure->Symptomatic seizures,Focal-onset seizure	4.354E-13

Figure 29. SAP score for phenotypes.

The right side of the panel, **Region Details**, displays information, as seen in **Figure 30**, on overlap with regions in the KnowledgeBase, as well as information on region overlap with external databases for both Pathogenic and Benign events (e.g., ClinGen, DECIPHER).

The **Knowledgebase Events** pane displays region overlap to Test Type matched CNV entries in the KB. For example, Constitutional Test Type samples will only display constitutional KB entries, while Oncology Test Type samples will only display oncology KB entries (for details on submitting to KB refer to pg. in user guide). The similarity of the event being reviewed to the KB event is displayed in the Similarity column of the **Knowledgebase Events** pane. This similarity calculation does not take CN event direction (loss/gain) into account.

The regions defined by the admin (in **Admin tab > Variant Details tab > BioDiscovery Provided Regions** folder) will be used to calculate and display similarity to **Pathogenic** and **Benign** regions in the region overlap pane. Only regions that have an overlap will be displayed.

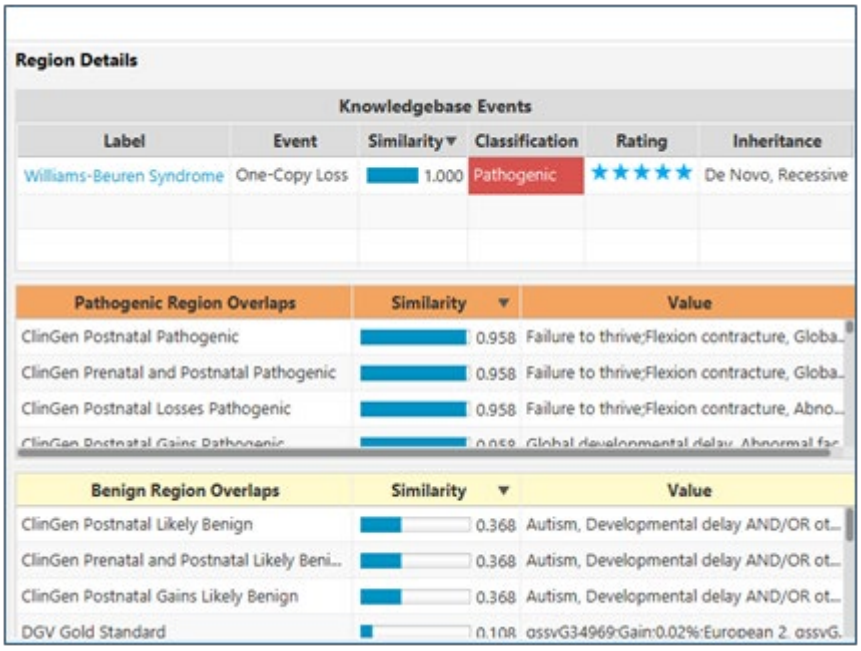


Figure 30. Regions that overlap with KB.

CNV Events – Oncology

Much of the content is the same for Oncology CNV events as that of constitutional events with the following notable display differences:

- Similar Previous Cases gauge
- The Mosaic status can be changed manually in Edit mode.

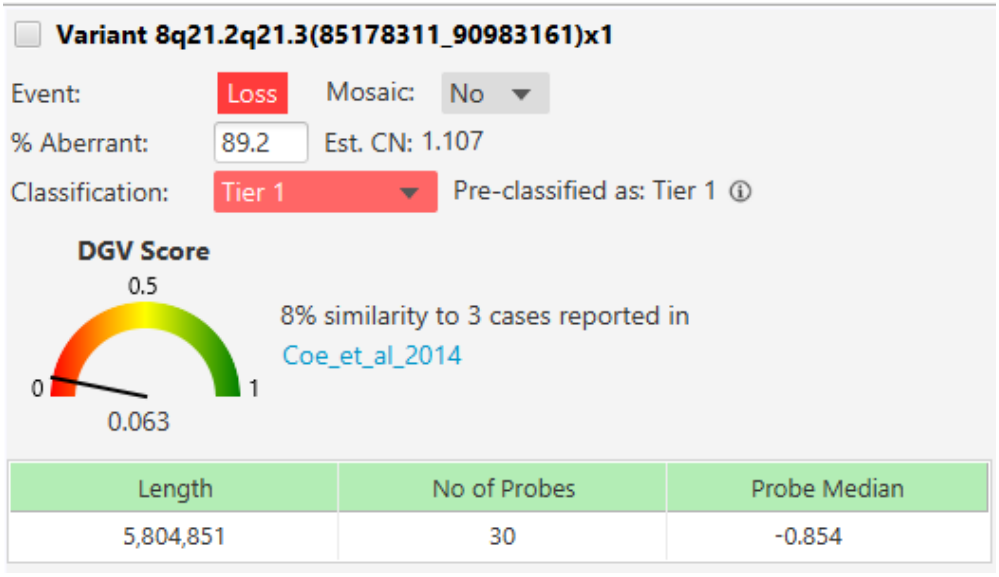


Figure 31. Oncology content differences for CNV.

- The **Region Details** pane displays content from the KB and Profiles, if relevant content is present, as in **Figure 32**.

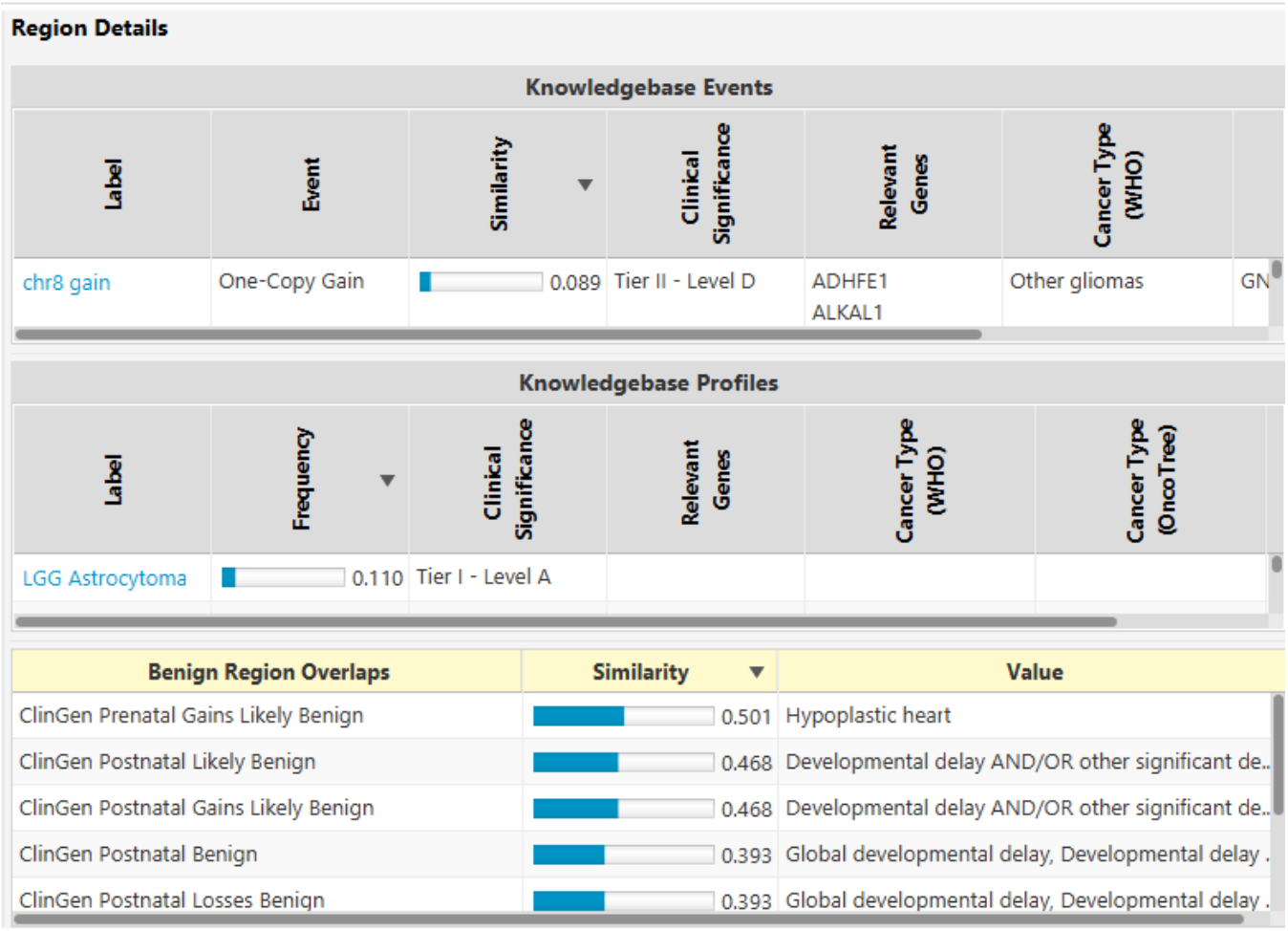


Figure 32. Oncology Region Details pane.

SeqVar Events

Some of the content for SeqVar events are the same as that for CNV events but there are additional features included, as seen in **Figure 33**.

The left-side panel, depicted in **Figure 34**, displays basic information about the variant (e.g., type, consequence, classification, ref/alt alleles, allele frequency, audit log). A dial representation of Similar Previous Cases is displayed. If the sample is part of a trio, Parent of Origin and Inheritance models are also displayed.

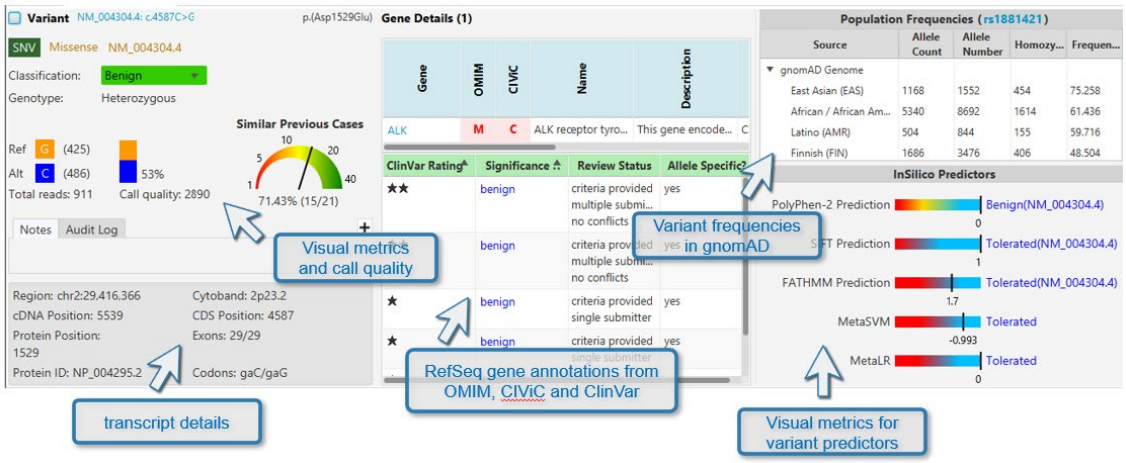


Figure 33. Additional features for SeqVar events.

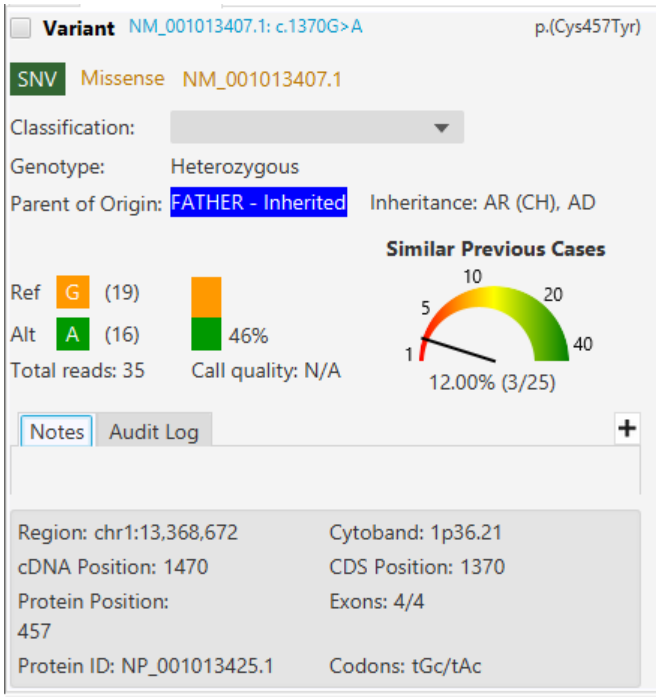


Figure 34. Left-side panel of a SeqVar event.

If ClinVar information is available, it is displayed in the **Gene Details** pane, as seen in **Figure 35**.

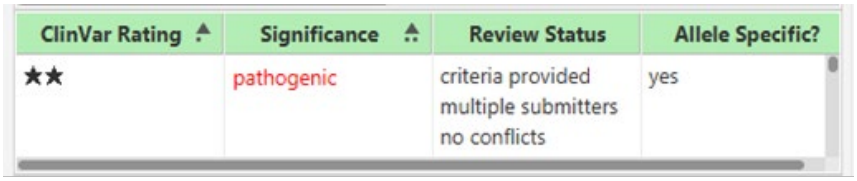
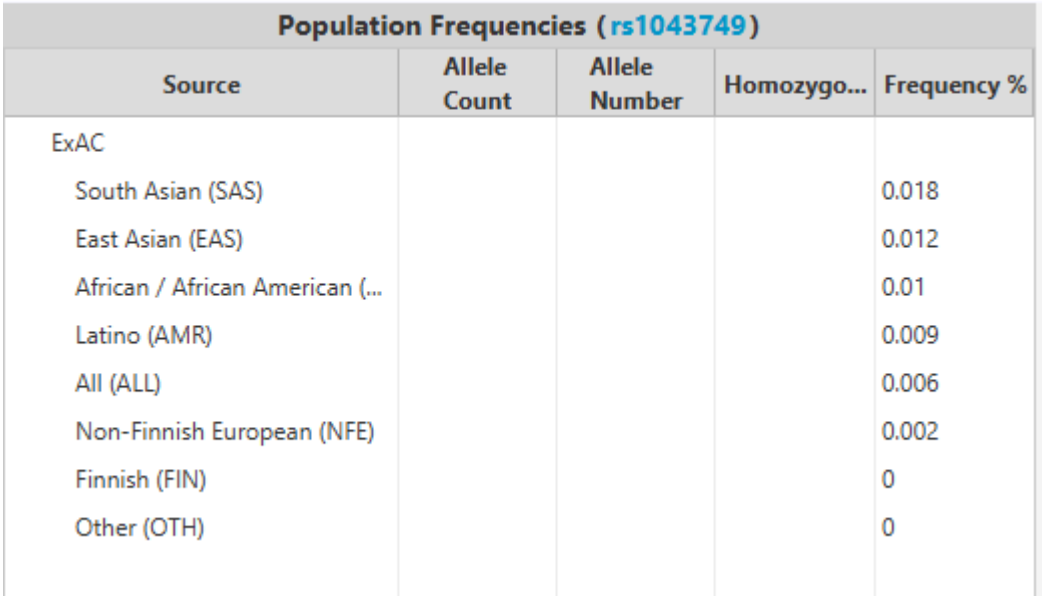


Figure 35. ClinVar rating for a SeqVar event.

If available, population frequencies are displayed in the right pane, as seen in **Figure 36**.



Population Frequencies (rs1043749)

Source	Allele Count	Allele Number	Homozygo...	Frequency %
ExAC				
South Asian (SAS)				0.018
East Asian (EAS)				0.012
African / African American (...				0.01
Latino (AMR)				0.009
All (ALL)				0.006
Non-Finnish European (NFE)				0.002
Finnish (FIN)				0
Other (OTH)				0

Figure 36. An example of Population Frequencies for a SeqVar event.

If one or more *in silico* predictors have values, the score and prediction is displayed in a graphical format in the right pane. The range goes from deleterious on the left (red) towards tolerated on the right (blue). The score is displayed under the bars with a hashmark. See **Figure 37**.

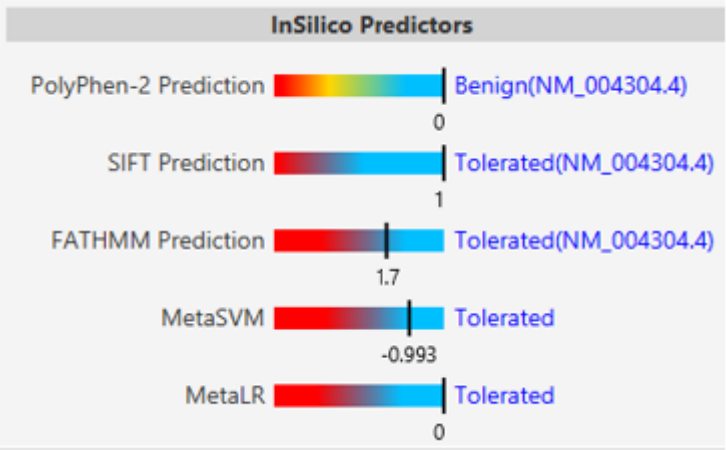
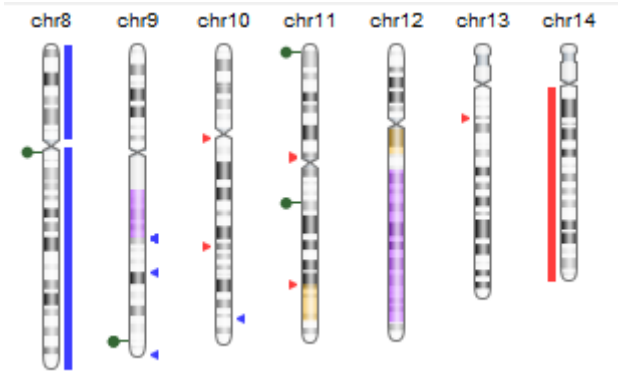


Figure 37. *In silico* predictors.

Tracks Tab

Initially, this section displays the ideogram at the top, showing events on and alongside of the chromosome diagrams. The type of events (CNV/allelic events, sequence variants) displayed depends on the sample type. Copy number and allelic events as colored bars next to and on the chromosome in the **Sample** tab. Shading on the chromosomes indicates AOH/allelic imbalance. Sequence variants (if present) are displayed as lines with dots (lollipops) jutting out from the chromosomes. Below the ideogram is a table listing events and other information about the regions, as seen in **Figure 38**.



CN and allelic events indicators:

Red = CN loss

Blue = CN gain

Gold = AOH

Purple = allelic imbalance

Sequence Variants indicators:

Different colors of the dots represent the type of sequence variant event:

Red – deletion **Green** – SNV **Gray** - indel

Blue - insertion **Purple** – MNV

Figure 38. Ideograms and indicators.

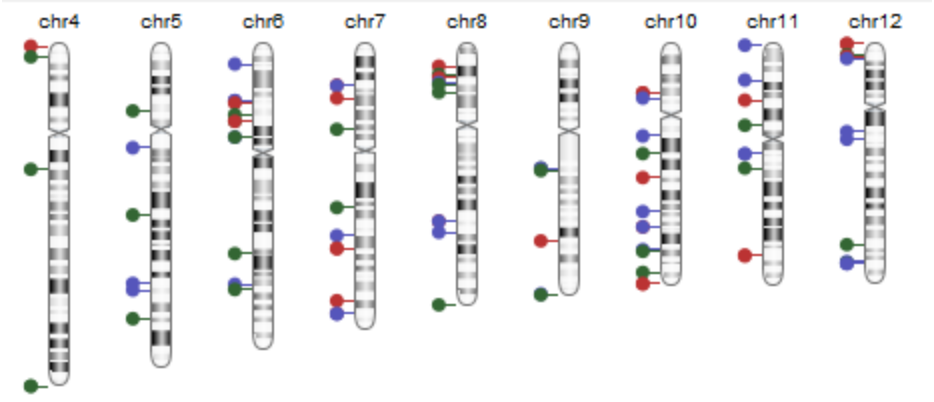


Figure 39. Example of a sample with only SeqVar events.

Depending on the sample type and modalities associated with a sample, both CNV/AOH and sequence variant events may not be available for a single sample.

Sample with only copy number and allelic events (often an array only sample type or an NGS sample for which copy number was estimated but no associated VCF or Nirvana JSON was loaded) is shown in **Figure 40** below.

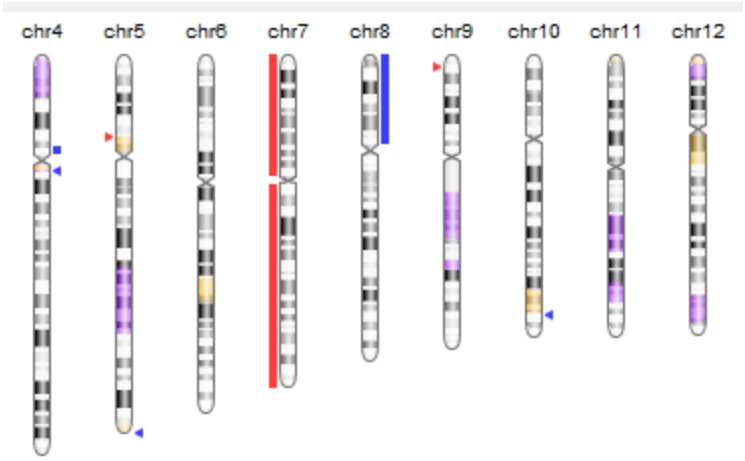


Figure 40. Only copy number and allelic events.

When an event or an ideogram is selected, the display zooms in on the selected area and displays an interactive genome browser with various tracks, as shown in Figure 41. Selecting a chromosome from the ideogram by clicking on the chromosome brings up the chromosome view with the events on that chromosome highlighted in the table.

Figure 41. Display zoom onto an interactive browser.

CN loss and gain events are displayed in solid red and blue colors respectively with deeper shades indicating big loss or high copy gain. Mosaic events are depicted with a textured fill. In **Figure 42**, the gain is mosaic (textured blue fill), but the loss (solid red) is not. Clicking on a table row will zoom in on the region in the panel above and place a dotted rectangle around the event.

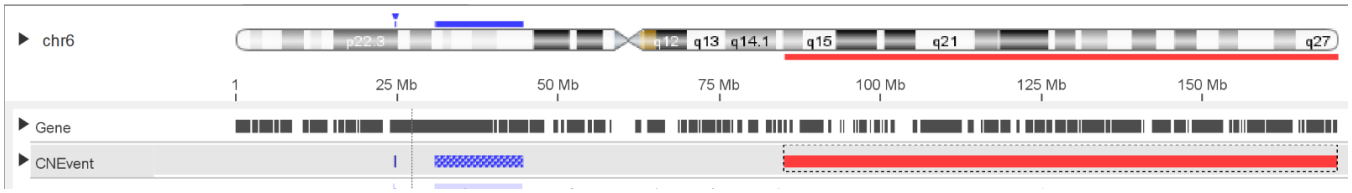


Figure 42. CN loss and gain, and mosaic events.

Clicking on the event in the browser will highlight the event row in the table, as seen in **Figure 43**.

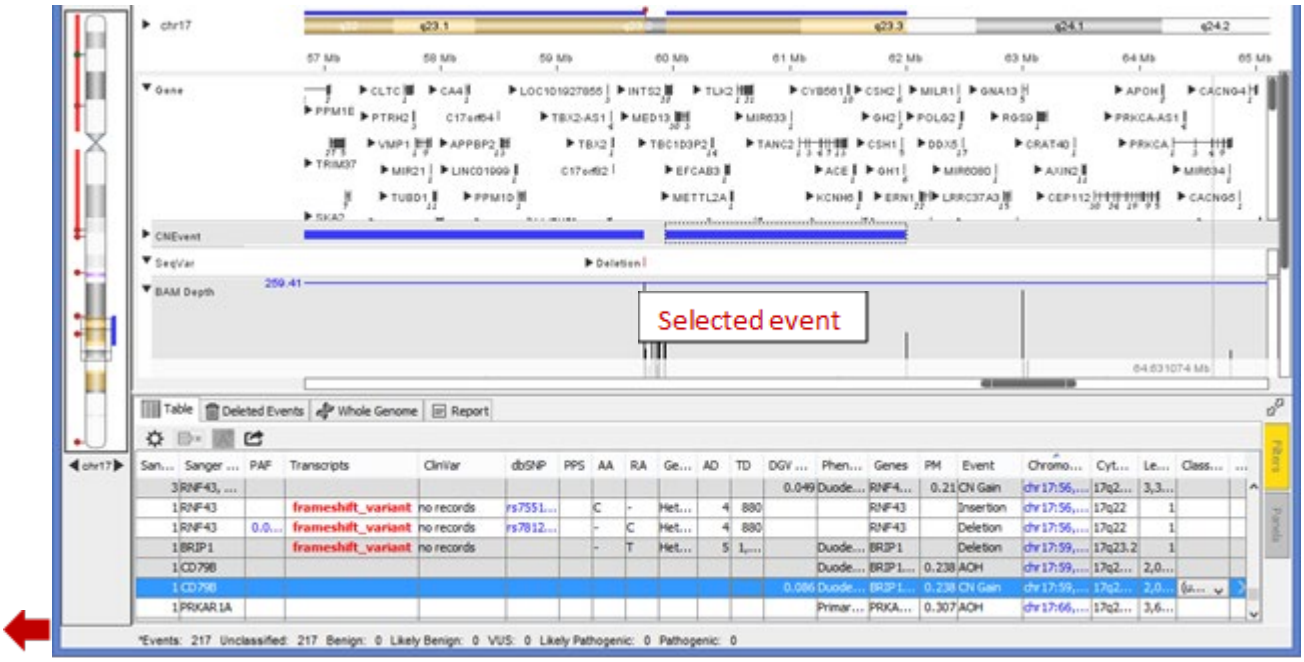


Figure 43. Highlighted event row.

When an individual chromosome is being displayed at the top, additional tracks are shown below. Clicking on the black arrow button next to the track name expands the track (arrow pointing down) to show details. **Figure 44** shows expanded **CN Event**, **BAM Depth**, and **Probes** tracks.

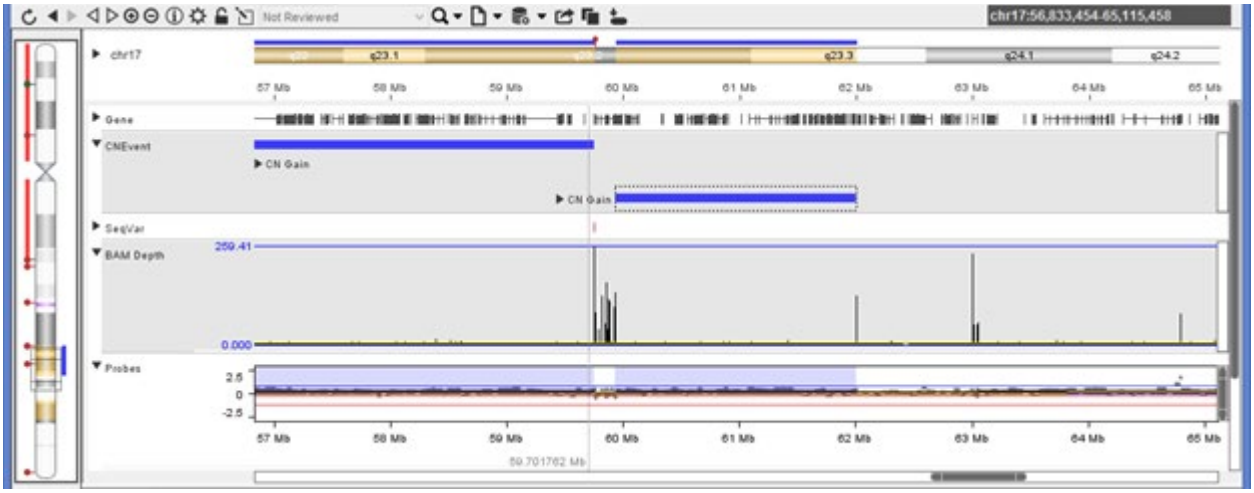


Figure 44. Expanded Probe tracks.

The height of a track can be adjusted by clicking on and dragging up or down the bottom boundary (horizontal gray line). **Figure 45** shows that the **BAM Depth** track was made shorter while the height of the **Probes** track was increased.

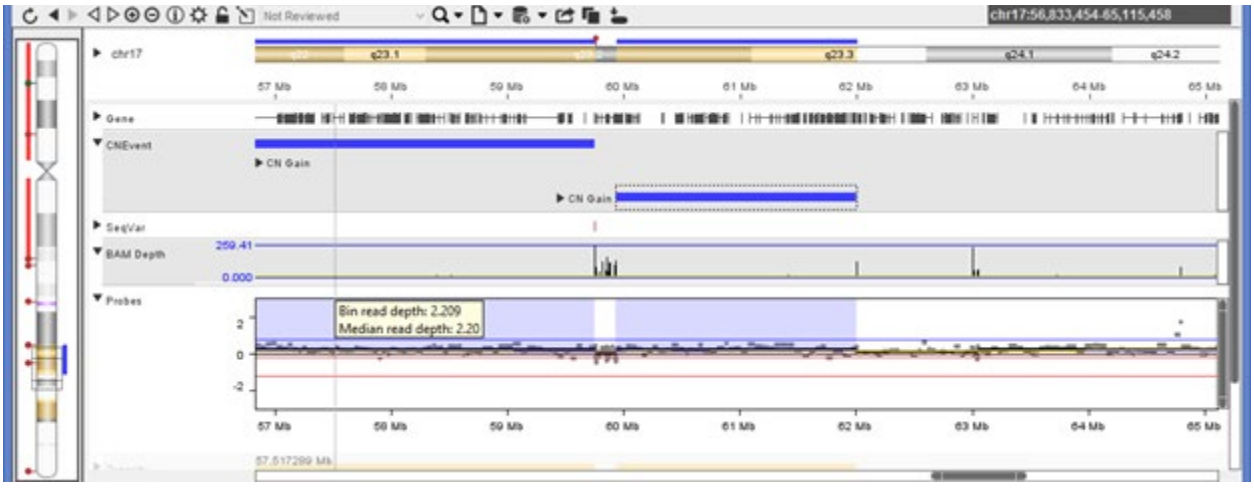


Figure 45. Track height.

The scale of the **Probes** track can be changed by dragging the bottom boundary. In **Figure 46**, the scale goes up to 2.5. After dragging the bottom boundary down, the plot height has increased and the scale now goes up to 3, as seen in **Figure 47**.

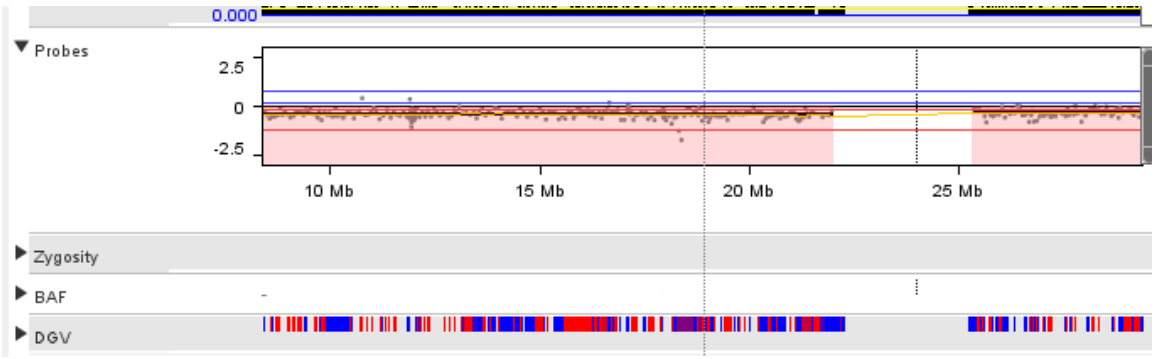


Figure 46. Probe scale.

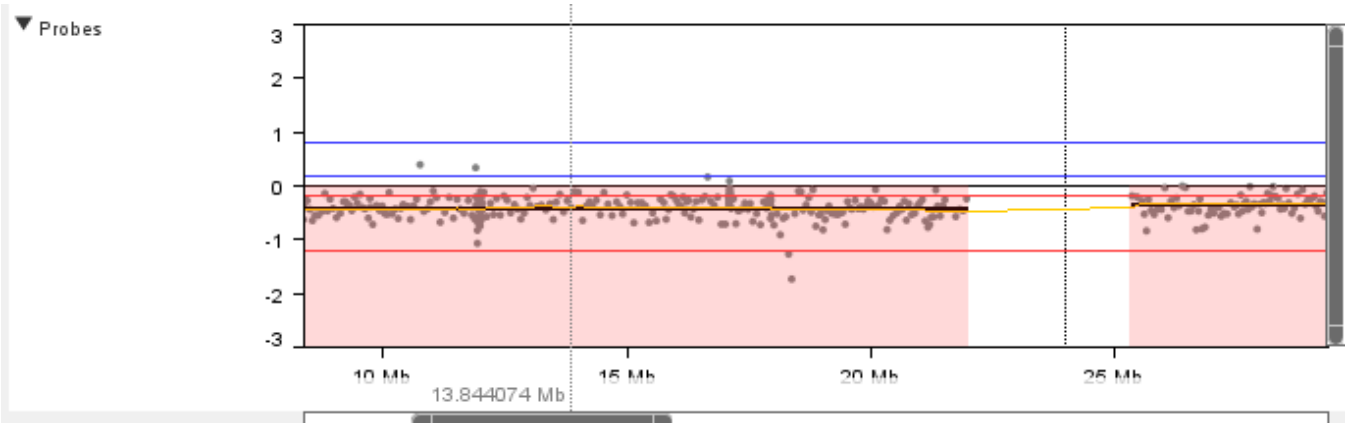


Figure 47. Increased plot height.

The **CN Event** track will show all the samples in the selected data types that overlap with the event in view. Darker red bars in the **CNEvent** track indicate homozygous copy loss and lighter red, single copy loss, as seen in **Figure 48**.

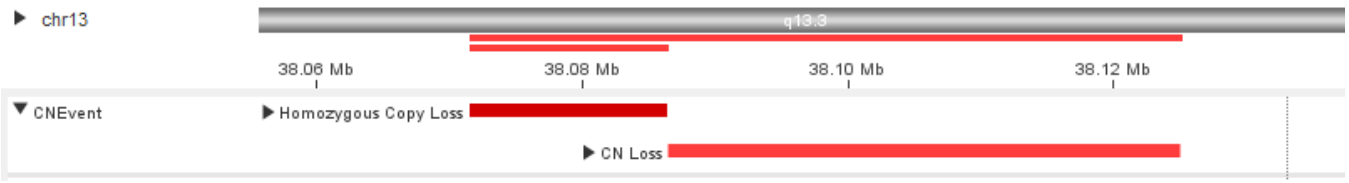


Figure 48. CN Event track.

Moving the mouse over each event shows information about that event, as seen in **Figure 49**, as well as the sample name in the yellow box.

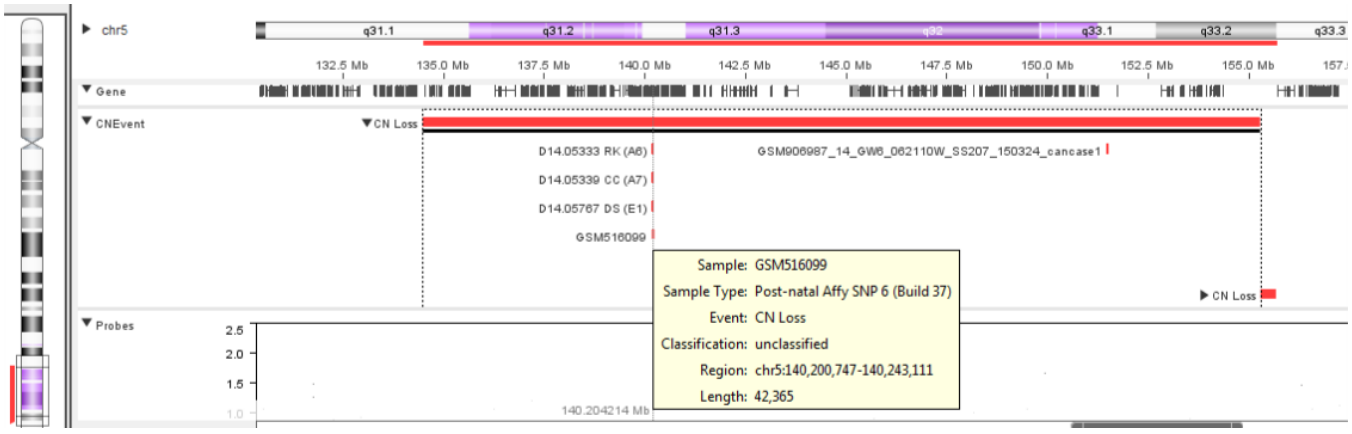


Figure 49. Event information.

Hovering over the event name and black triangle displays how many cases with similar events (including the current case) are present in the database, shown in **Figure 50**.

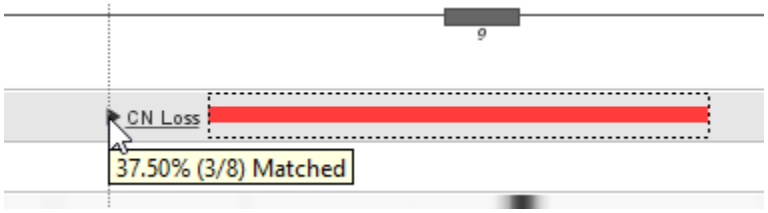


Figure 50. Cases with similar events.

Right clicking on an event will bring up a menu, shown in **Figure 51**, allowing additional functions (functions will be active only if the user has permissions to perform those). If sample editing is on, the menu will allow classifying a call as well as deleting or modifying it.

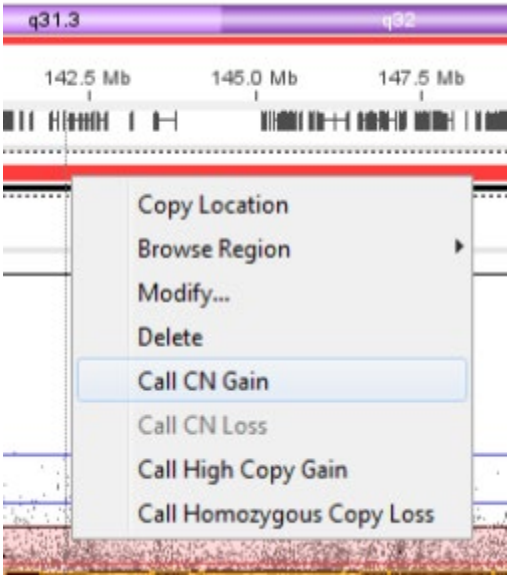


Figure 51. Additional function menu.

To show individual probes relating to an event, open the probe track. Moving the mouse over individual probes will display information about the probe including the log ratio.

Depending on network speed, display of probes may take some time especially on slower networks. It is possible that the probes may not be visible immediately upon opening a new sample that has previously not been opened on the computer being used for sample review. A spinning circular icon on the bottom right, seen in **Figure 52**, indicates that probes are being loaded.

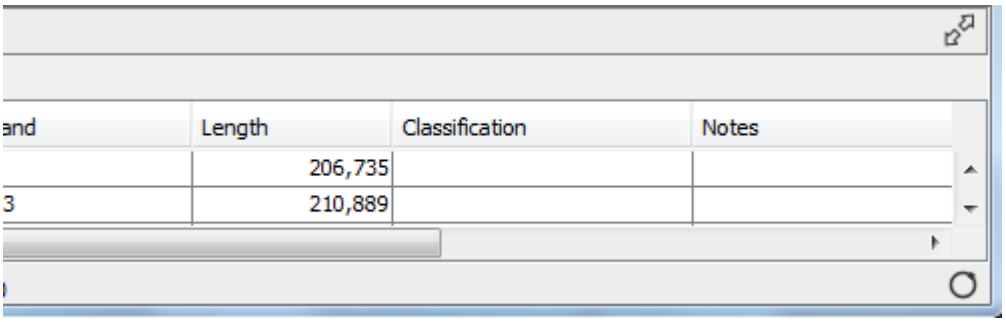


Figure 52. A spinning circular icon on the bottom right.

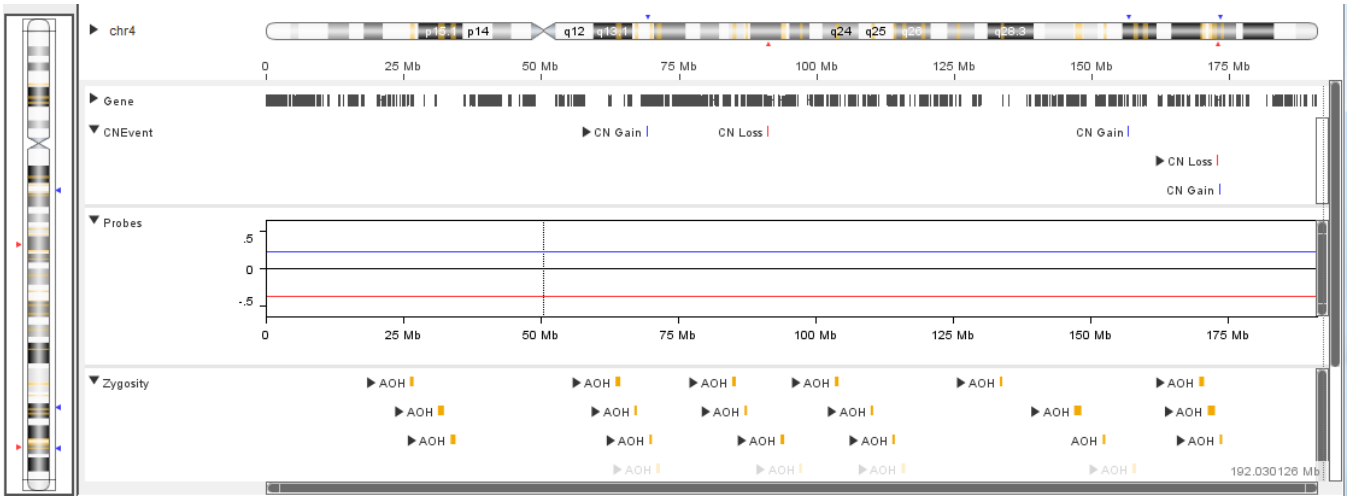


Figure 53. Probes have not yet been fully downloaded; therefore, no probes are visible.

Once probes have been fully downloaded, they will be displayed in the **Probes** track, as seen in **Figure 54**.

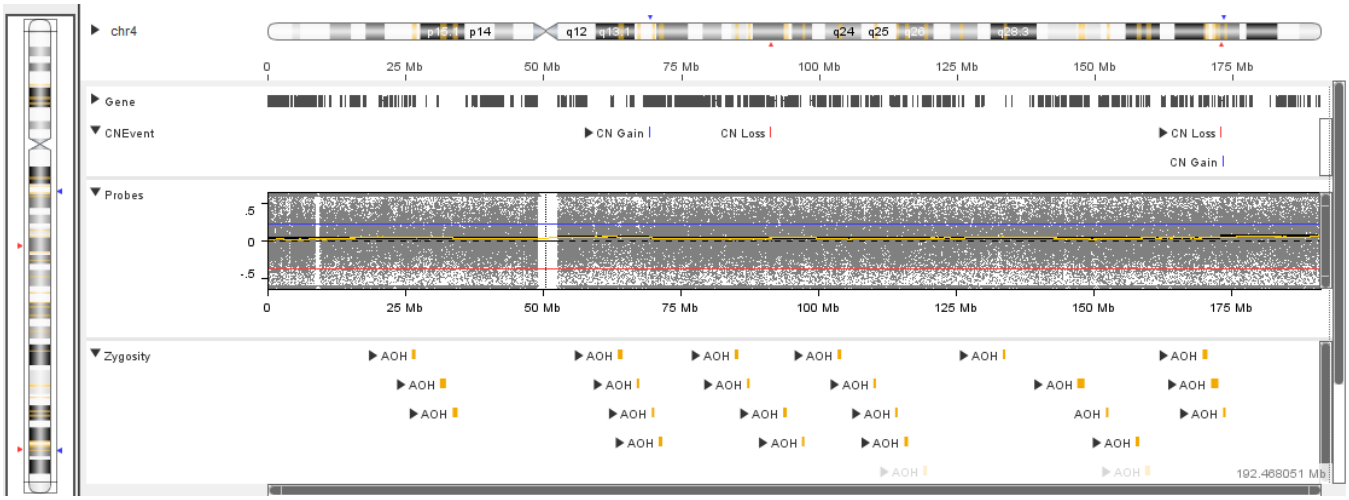


Figure 54. Fully downloaded **Probes** track.

Sequence Track

Zooming in on an event to the base level reveals the sequence track displaying the individual bases in the reference genome. Nucleotides are color coded: A – green, C – blue, G – yellow, T – red, as shown in **Figure 55**.

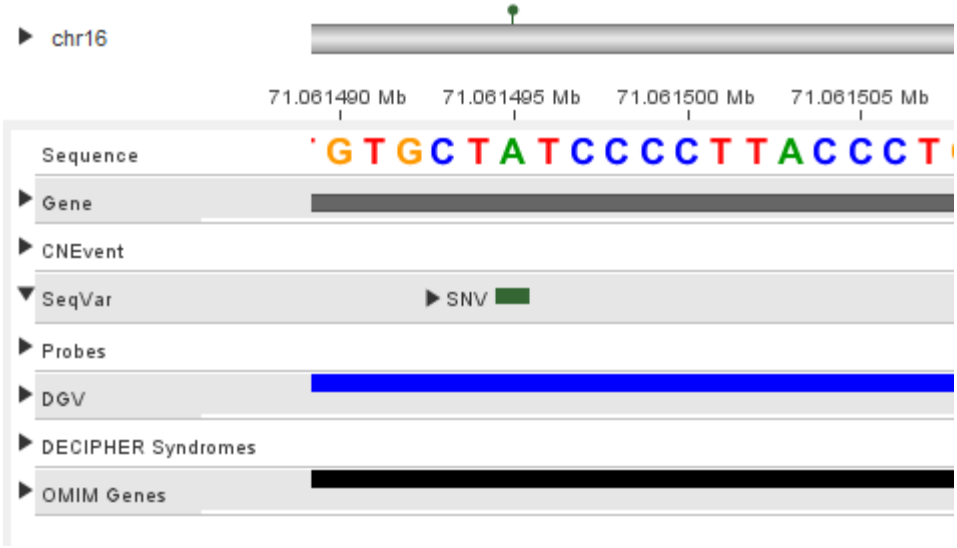


Figure 55. Color-coded nucleotides.

Gene Track

The Gene track displays genes from RefSeq within the region in the current view. Genes are collapsed and compacted into black bars when zoomed out, as shown in **Figure 56**. Upon zooming in to a sufficient level, gene symbols begin to appear. Further zooming in will reveal the gene structure with exon indicators. Genes are displayed as a union of all transcripts in the compact state with hash marks indicating exons. Exon numbering is based on the transcript selected and by default this is the canonical transcript (from **Nirvana** annotator which obtains canonical information from VEP).

Figure 56. Gene track.

Each gene can be expanded (via the black triangular icon next to the gene symbol) to reveal all transcripts, shown in **Figure 57**. The transcript shown in black is the one chosen for numbering exons in compact view.

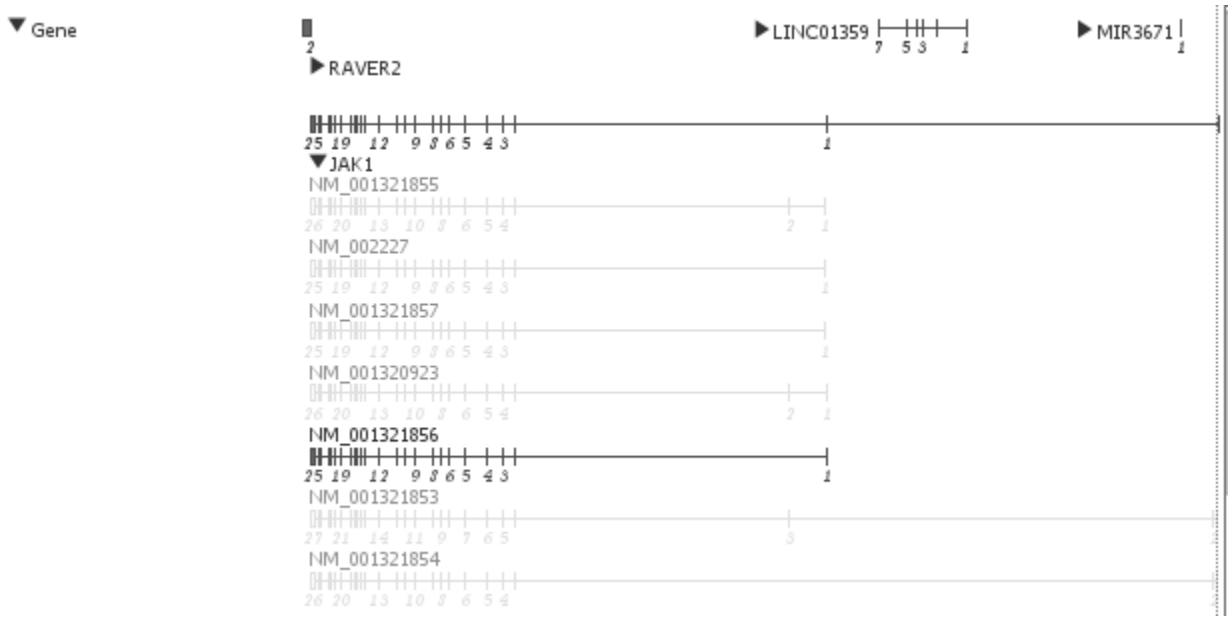


Figure 57. Expanded gene transcripts.

To select a different transcript to use for numbering exons, right click on the transcript ID and choose a different item from the list, as shown in **Figure 58**. Notice that the exon numbering for the gene has now changed and is based on the selected transcript (NM_001321856), shown in **Figure 59**. The selected transcript is displayed in black while the others are in gray.

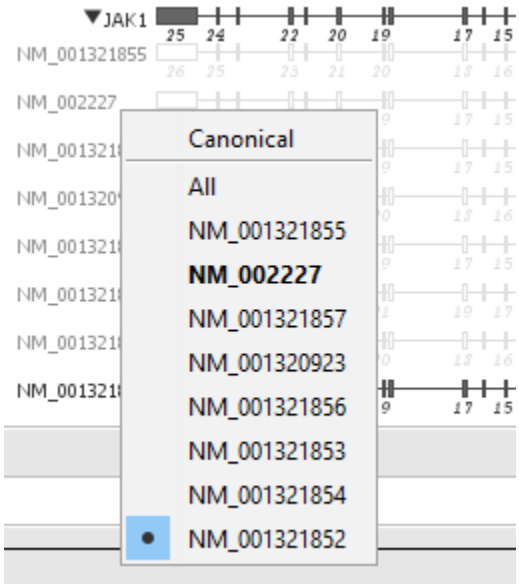


Figure 58. Selecting a different transcript.

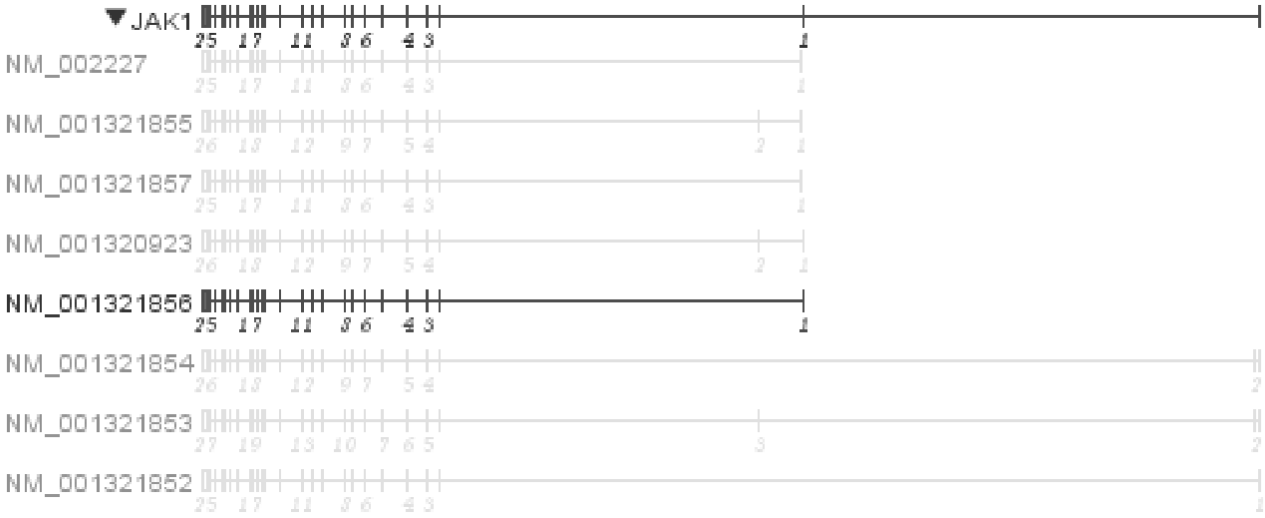


Figure 59. A different, selected transcript.

The list of transcripts now has the blue box with dot next to the one selected for exon display (NM_003121856), shown in **Figure 60**. The canonical transcript for this gene is still displayed in bold.

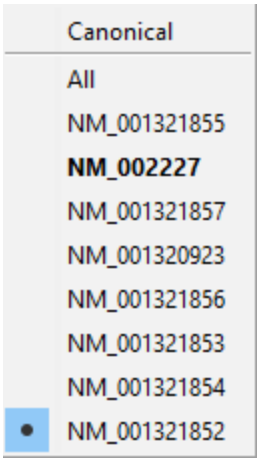


Figure 60. List of transcripts.

SeqVar Track

Like the **CN Event** track for copy number, the **SeqVar** track displays the sequence variant events as well as any overlapping events from past cases, as displayed in **Figure 61**.

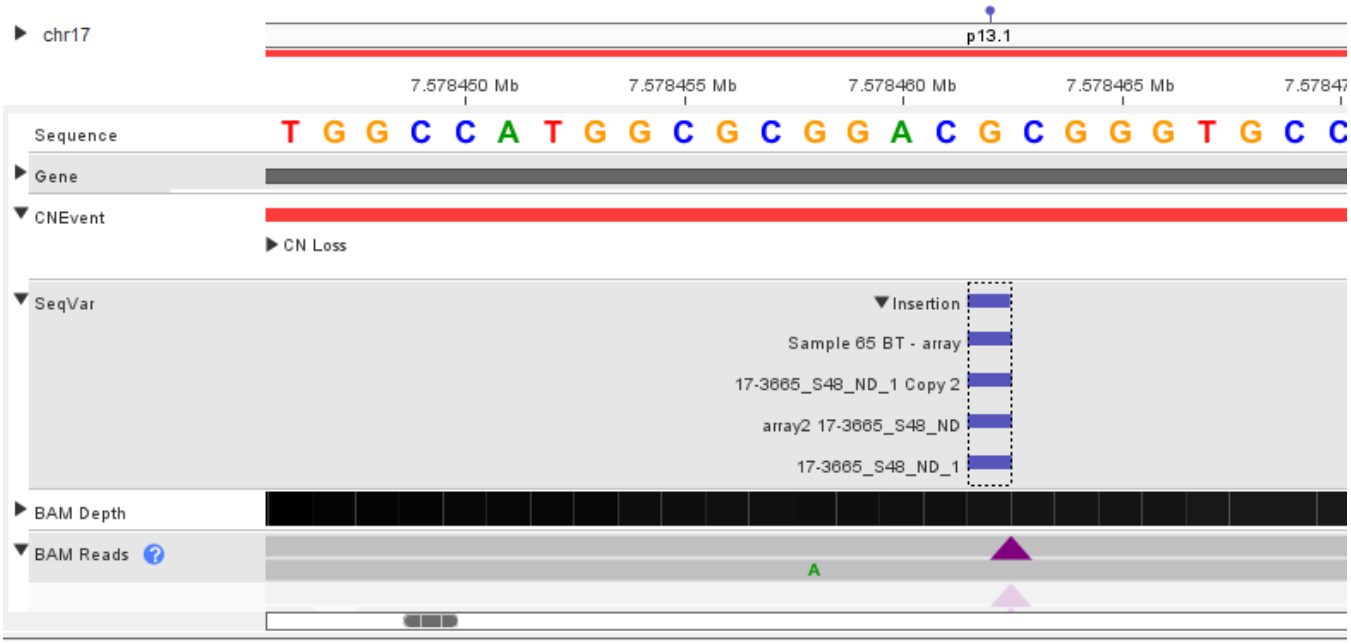


Figure 61. SeqVar track.

BAM Depth Track

The **BAM Depth** track displays the read depth obtained from BAM files. In the collapsed mode, the track will display gray and black lines in the read areas, shown in **Figure 62**, the darker the color, the higher the read depth in that region. Hovering over the reads will display the bin read depth and median read depth.

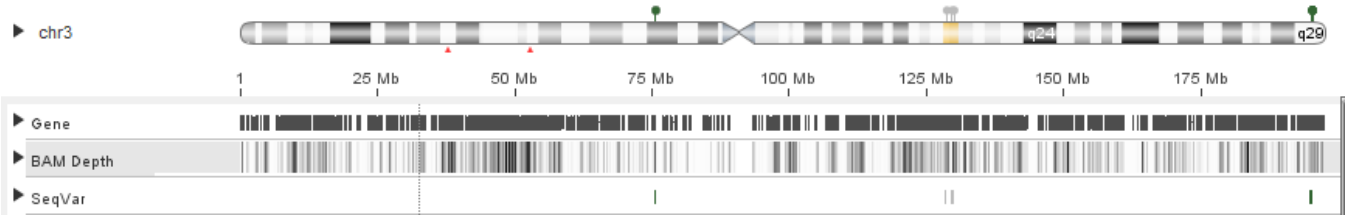


Figure 62. BAM Depth track.

In the expanded track mode, vertical lines will be displayed as the height correlating to the bin read depth. The Y axis is a dynamic sigmoid scaling of the read depth with horizontal blue lines as tick marks between the maximum and 0, seen in **Figure 63**. The maximum scale represents the largest bin read depth within the range displayed in the window. As one zooms in and out, the Y scale changes based on the maximum read depth of the region displayed. The horizontal yellow line marks the median read depth within the chromosomal range displayed in the window. Hovering over the plot will display the bin read depth at that location.

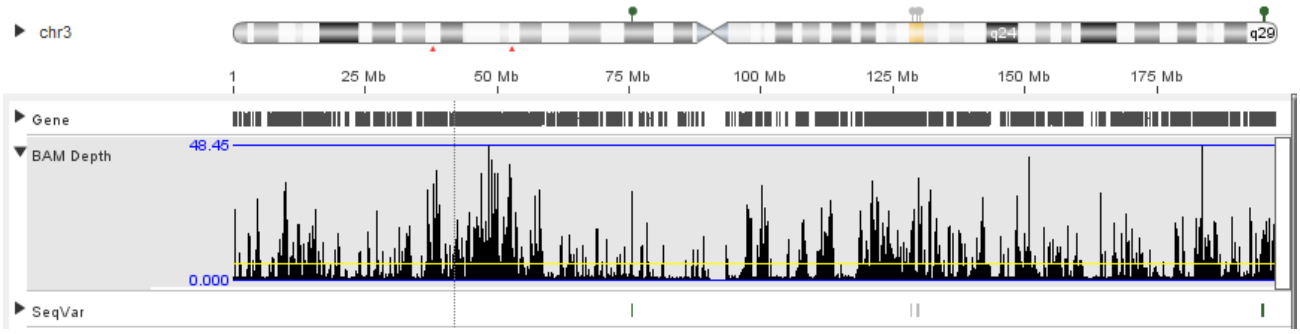


Figure 63. Expanded Track mode.

Zooming in sufficiently will make visible the **BAM Reads** track.

Figure 64 shows how the BAM Read depth displays a high density of reads which exactly correlates with exons in the targeted panel sample.

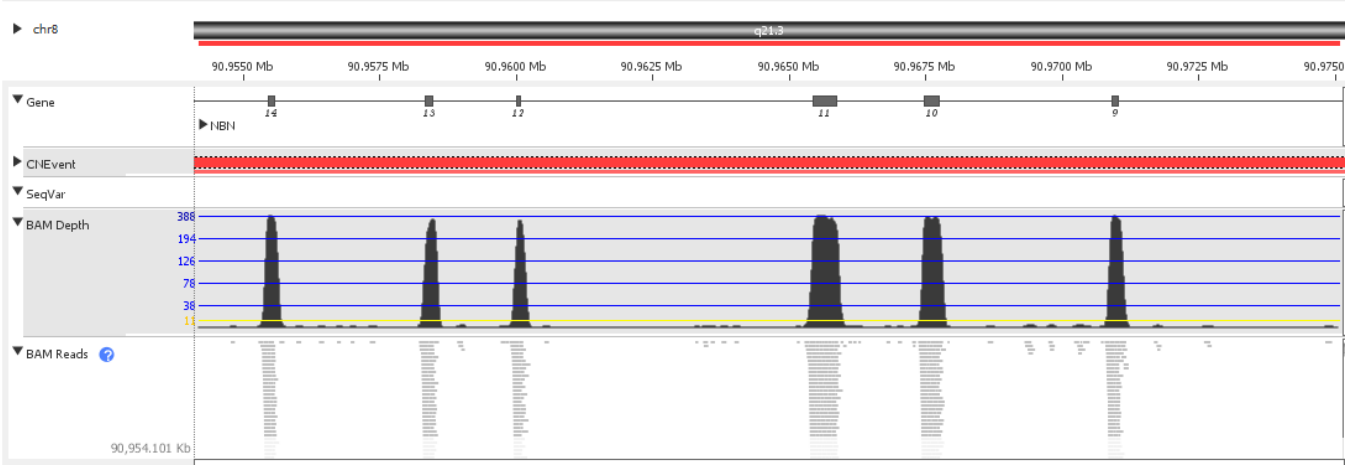


Figure 64. BAM Reads track.

Zooming in even further will make visible the **Sequence** track, shown in **Figure 65**, displaying individual bases in the reference genome:

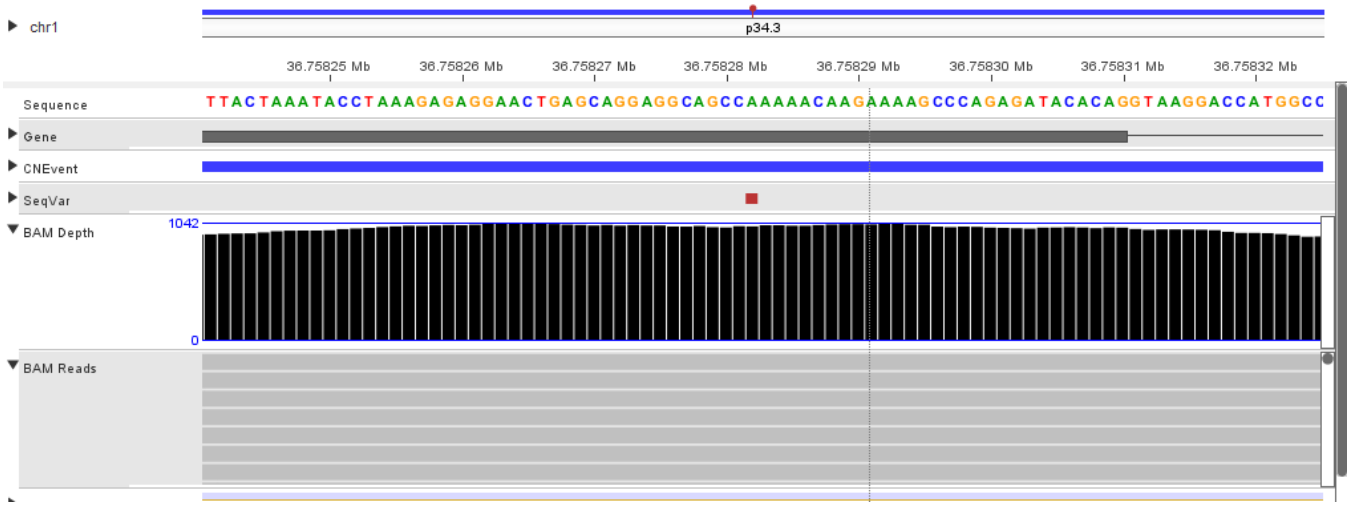


Figure 65. Sequence track.

The height and color coding of the bars displays the numbers of each nucleotide found in that location. In **Figure 66**, the bars are extremely short over a two base pair region indicating a two-nucleotide base pair deletion. Hovering over a location shows the number of each nucleotide counted at that position.

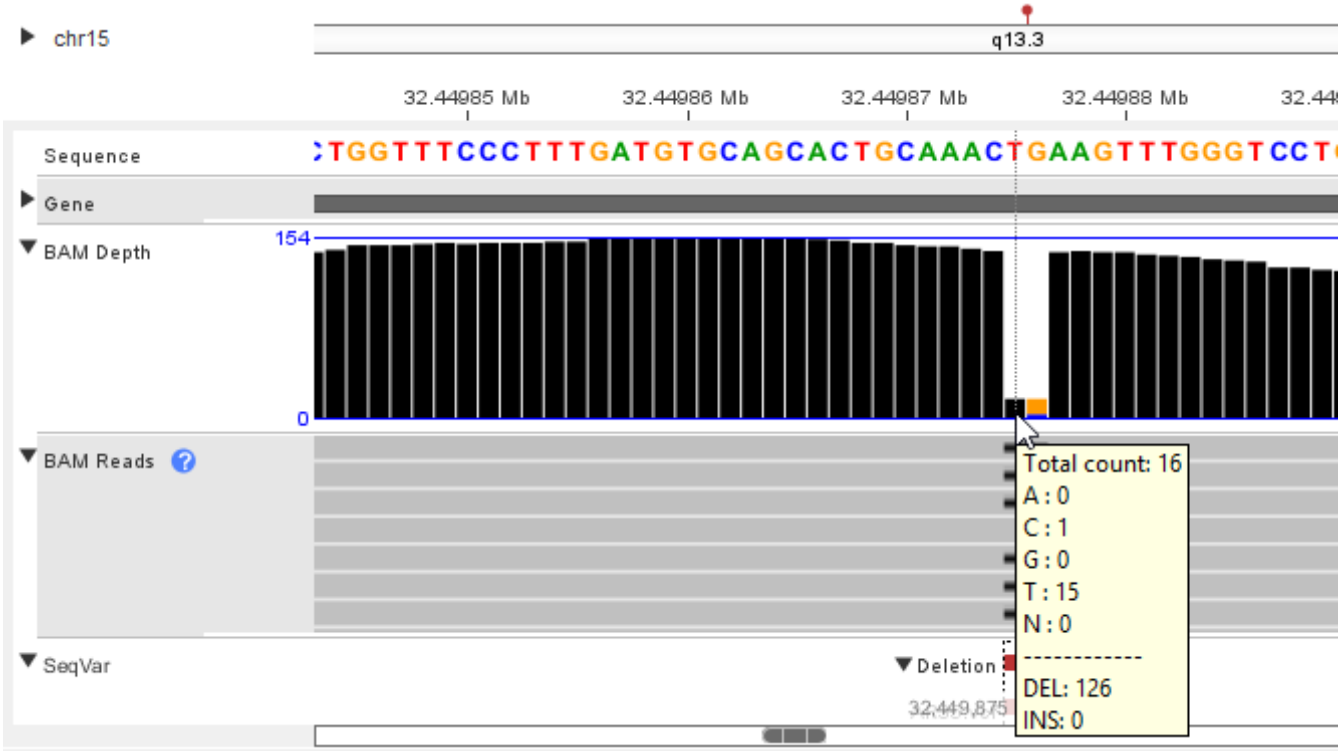


Figure 66. A two-nucleotide base pair deletion.

BAM READS

The **BAM Reads** track is visible when zoomed in sufficiently on the browser. Hovering over a read displays the read quality, as shown in **Figure 67**.

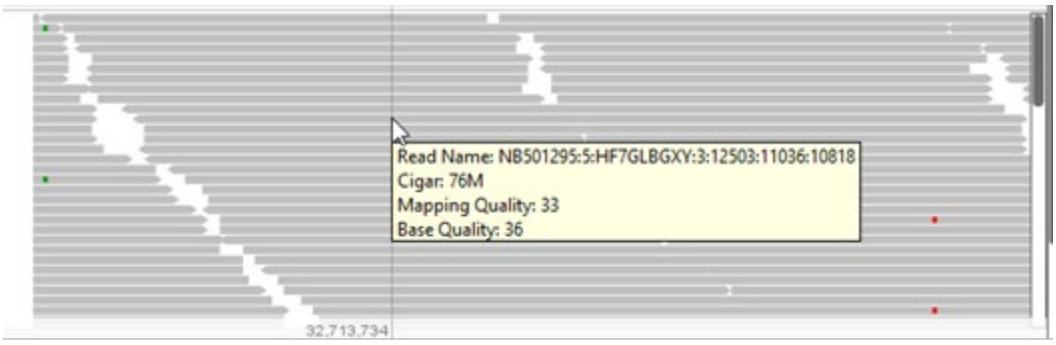


Figure 67. Read quality.

Color coding of individual reads is seen in **Figure 68**.

- **Darker gray:** standard mapped reads.
- **Light gray:** read mapping is a secondary or supplementary alignment or read is a paired read and the alignments are not what would be expected of a proper pair or an unmapped pair mate.
- **Red:** read is a paired read but its mate is mapped to a different chromosome (the read pair spans a translocation breakpoint, or the read(s) were mapped incorrectly).
- **Blue:** reads (paired) are mapped in the wrong orientation relative to each other (the read pair spans an inversion breakpoint, or the read(s) were mapped incorrectly).
- **Green:** read pair spans a deletion or there is a mapping error (spacing between the reads is more than 10,000 nucleotides).

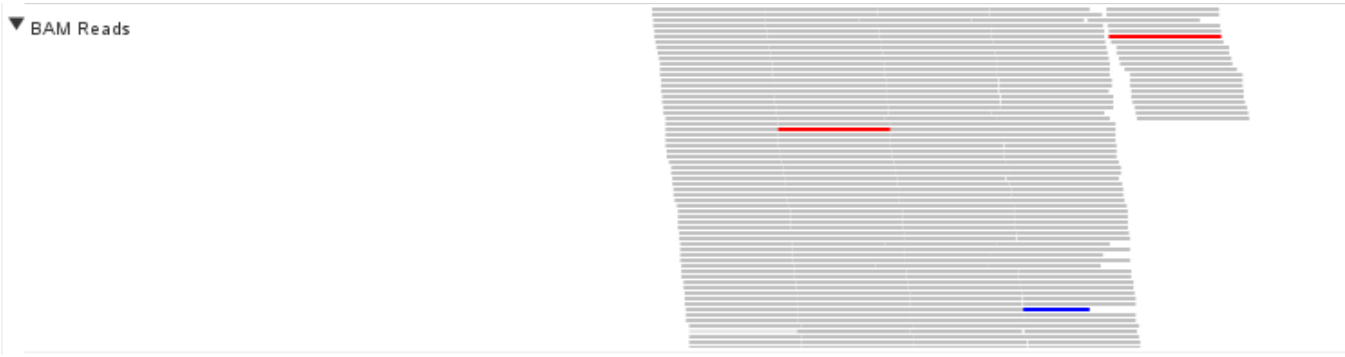


Figure 68. Individual read color coding.

Nucleotides in the reads: If the bases in the reads match the reference sequence, the base letter is not shown. Mismatches in reads are shown as the mismatch letter. Deletions are shown as gaps with a black line running through them. A two base pair deletion is displayed in **Figure 69**. Insertions are shown with purple triangles, as seen in **Figure 70**. The thymine and guanine bases have been deleted in several of the reads.

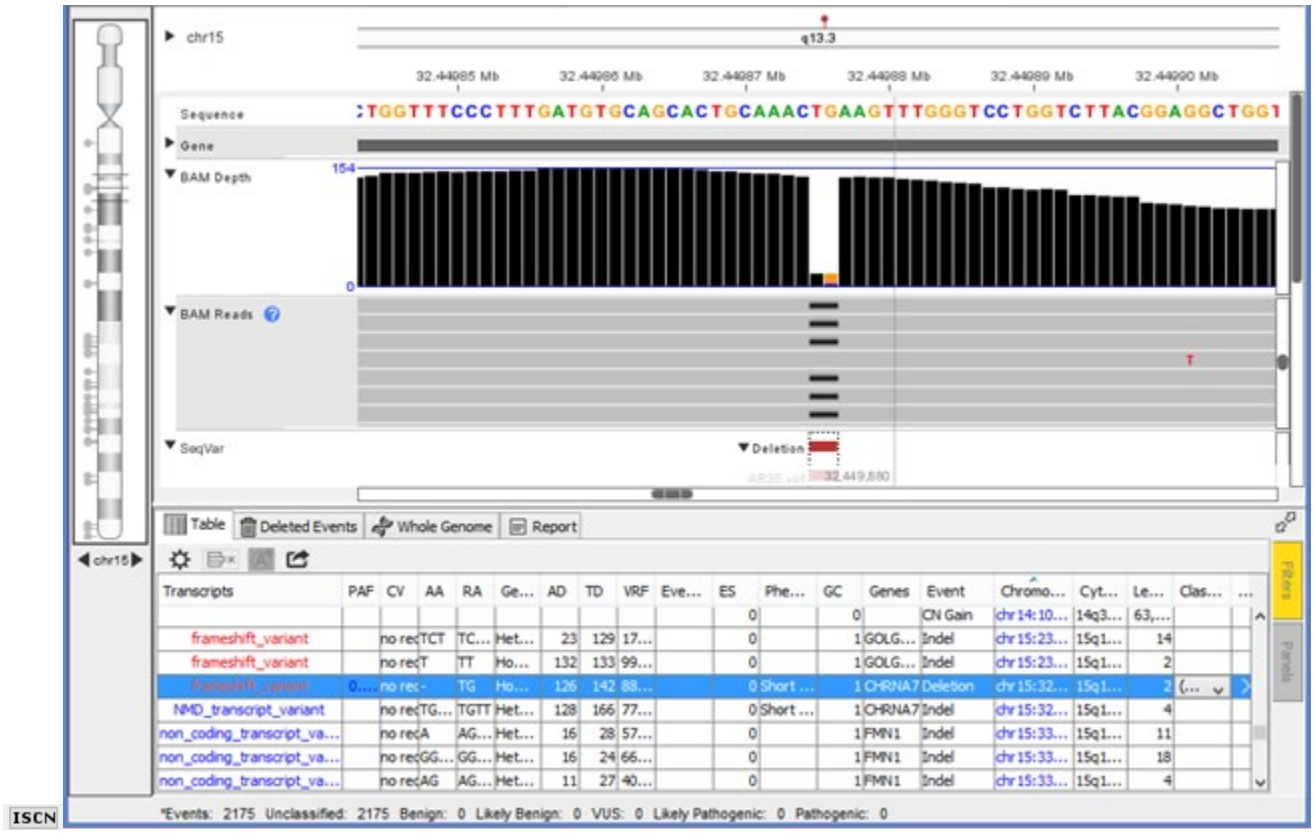


Figure 69. A two base pair deletion.

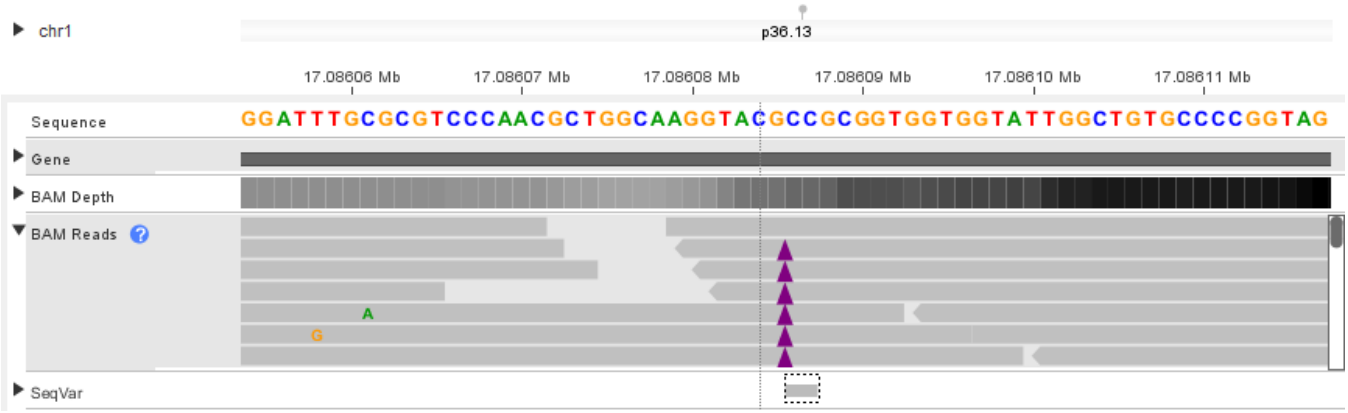


Figure 70. Insertion depicted by purple triangles.

Graphical Display Navigation and Toolbar

Across the top of the interface, above the ideogram, lies the toolbar along with a search field, represented by **Figure 71**.

Figure 71. The toolbar.

The filled triangle buttons are like web browser navigation buttons. They allow the user to go back to the previous page (or previous action performed) or forward to the next page (or subsequent action performed). By zooming in

on a region and then clicking the back button, the view will zoom back out to the original view. By clicking on the forward button, the zoomed in view will be brought back.

The non-filled triangles step backwards or forwards to the prior or subsequent event listed in the table. The view at the top will be zoomed in on the event and the table will have the event row highlighted. The plus and minus buttons allow zooming in and out. One can also zoom in by clicking and dragging on the ideogram.

Updating Sample Information

The **Sample Information** button brings up another window to show items and details about the sample such as its attributes, setting used to process the samples, sample name, QC fields, processing settings, and more. See **Figure 72**.

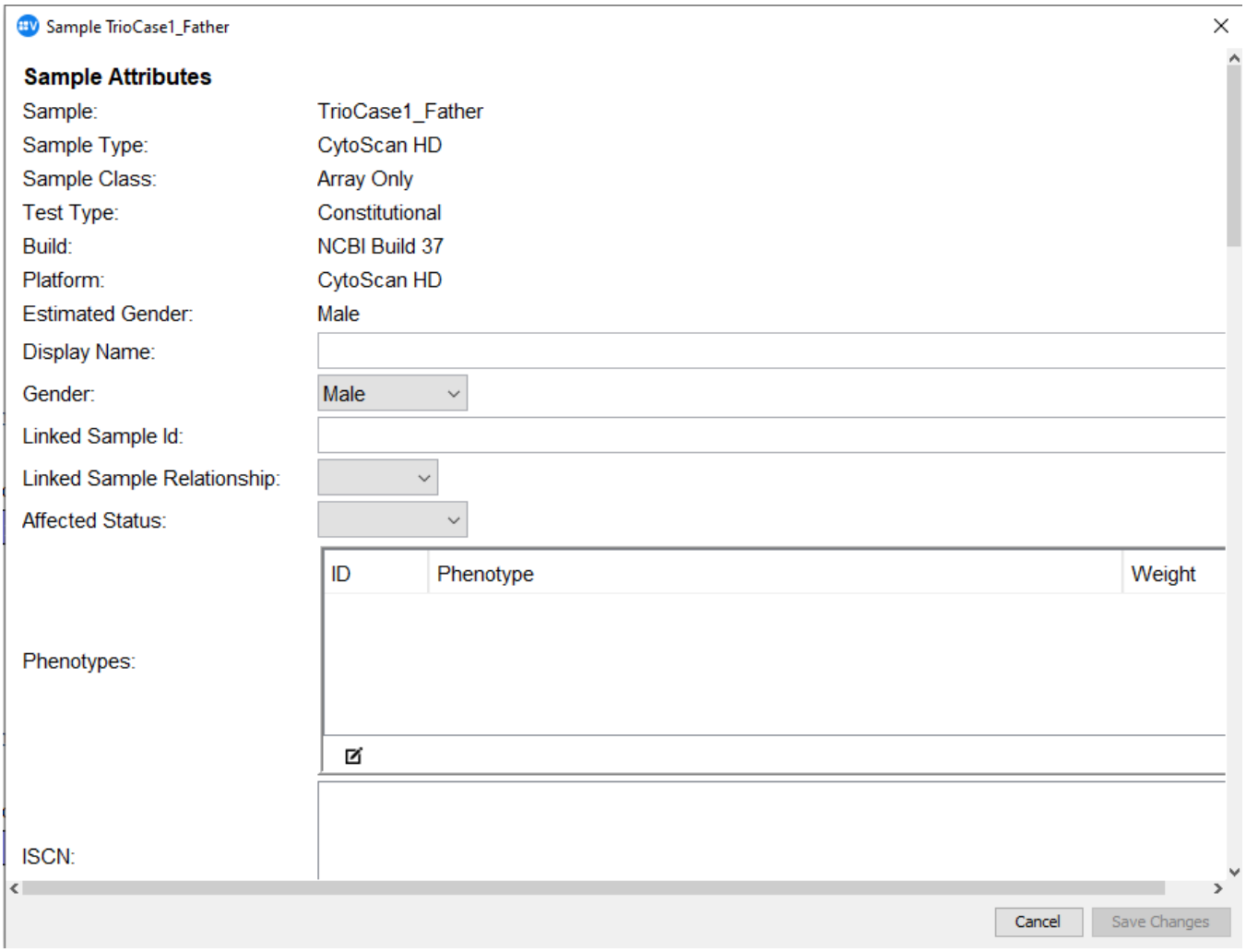


Figure 72. Updating sample information.

If a Decision Tree was used with this sample, the name(s) will be displayed in the **Sample Information** window. One or two decision tree names will be displayed depending on which DT was applied and to what data. When

adding SeqVar to an existing CNV sample, if a different DT is used than the one applied to CNV, the DT names will be displayed with CNV Auto pre-classification and SeqVar Auto pre-classification. If the same DT is used, then only one DT name is displayed with the label **Auto pre-classification**.

Users are also able to edit and specify important information about the sample such as gender, whether the sample is part of a linked analysis, and phenotype (please review the section *Creating and Visualizing Related Samples/Trio Analysis* for details on linked samples). To edit the gender of the sample, select the correct gender from the dropdown menu. It is important to specify the gender if known, as aberration detection and event classification is influenced by gender. Incorrect or no gender selection may give a misleading result. The phenotype of the sample can be edited via the **EDIT** button. A new window appears listing human phenotype ontologies (HPO), shown in **Figure 73**.

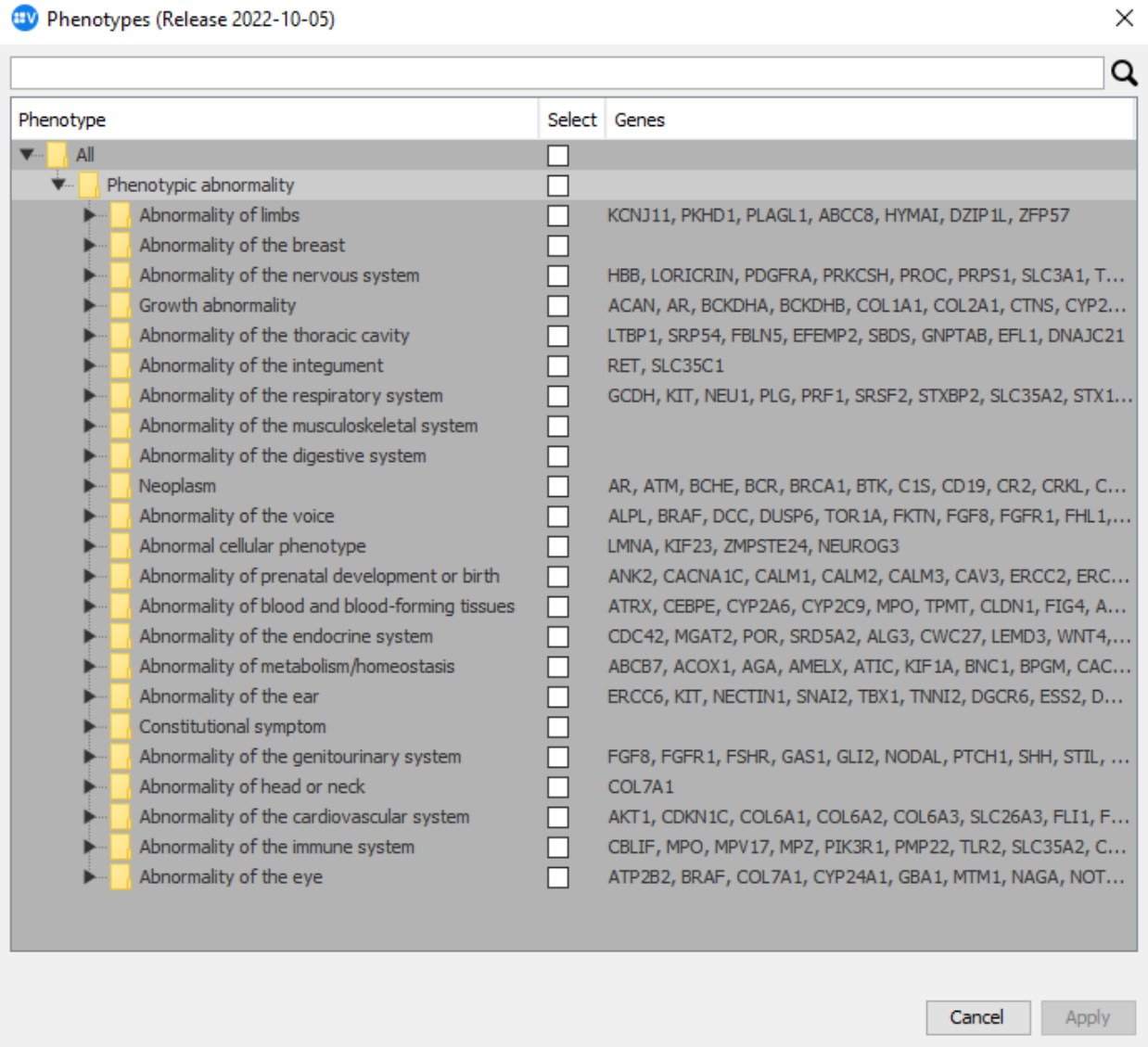


Figure 73. Human Phenotype Ontology list.

Typing key words that describe the sample's phenotype, e.g., speech delay, heart malformation, autism, allows users to search for the most appropriate HPO descriptor. Use the check boxes to select the required phenotype and then click **Apply**. It is important to specify the phenotype correctly because this information can be used to aid event classification.

Capture Bias

For panel and exome data a Capture Bias score is displayed (see section on Capture Bias in the System Administration Guide for details). Only users with processing privileges may re-process the sample to see if an alternate processing setting provides improved call quality (see **Figure 74** for an example of a poor score). The recommendation is to re-sequence samples with poor scores. **Figure 75** is an example of reprocessing.

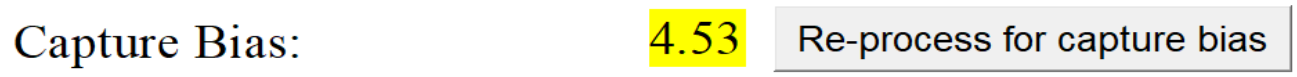


Figure 74. Sample with poor score.

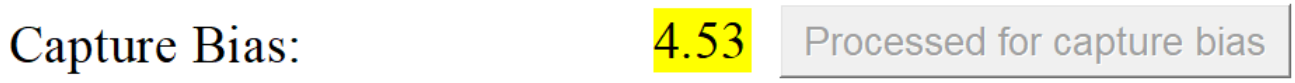


Figure 75. Sample with poor score that was re-processed.

Linked Nirvana Annotator and Data Source Versioning

If samples were processed using the linked Nirvana Annotator, the annotator version, data version and versions of the individual data sources will be displayed as seen in **Figure 76**.

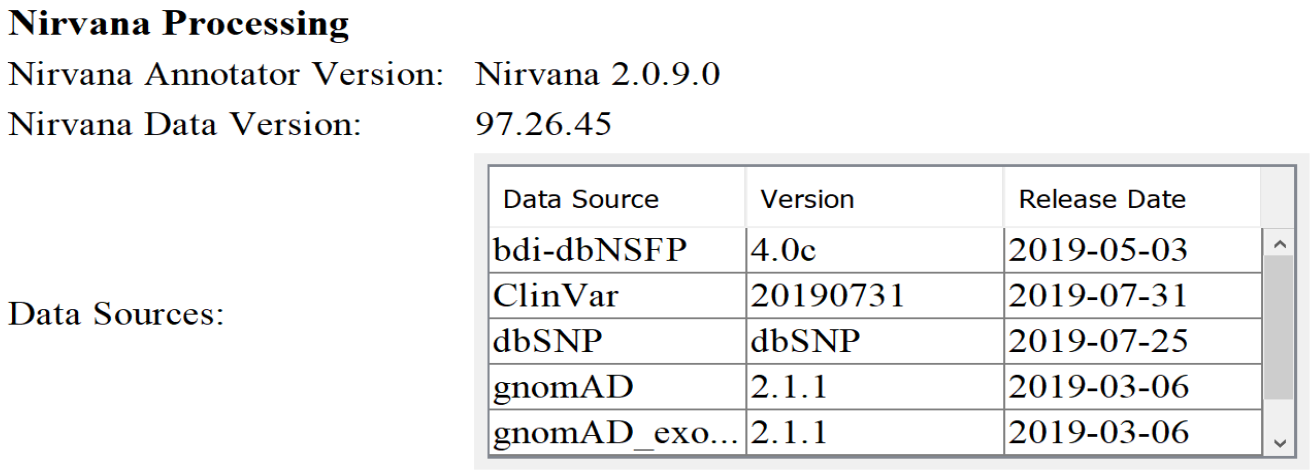


Figure 76. Samples processed using Nirvana.

The Nirvana Data Version numbers (97.26.45) correspond to the following, in order:

- VEP cache from which the data was obtained (97 in the example above)
- Cache version (26 in the example above)
- SA version (45 in the example above)

Modifying Track Display

The **Table and Track Preferences** button allows users to choose preferences as to how data is to be viewed and displayed. To change which tracks are displayed and in what order, first ensure that the **Tracks** tab is selected. The **Table** tab is explained in the **Data Table** section.

Click on the appropriate check box to display a track in the **Sample** window, shown in **Figure 77**.

NOTE: to see all the available tracks click on the plus signs next to the folder symbols. The width of the first column in the **Tracks** and **Table** tabs is adjustable so that the full track and column name is visible. Click on the right edge of the header row (shown in **Figure 78**) and drag left/right to make the column narrower/wider.

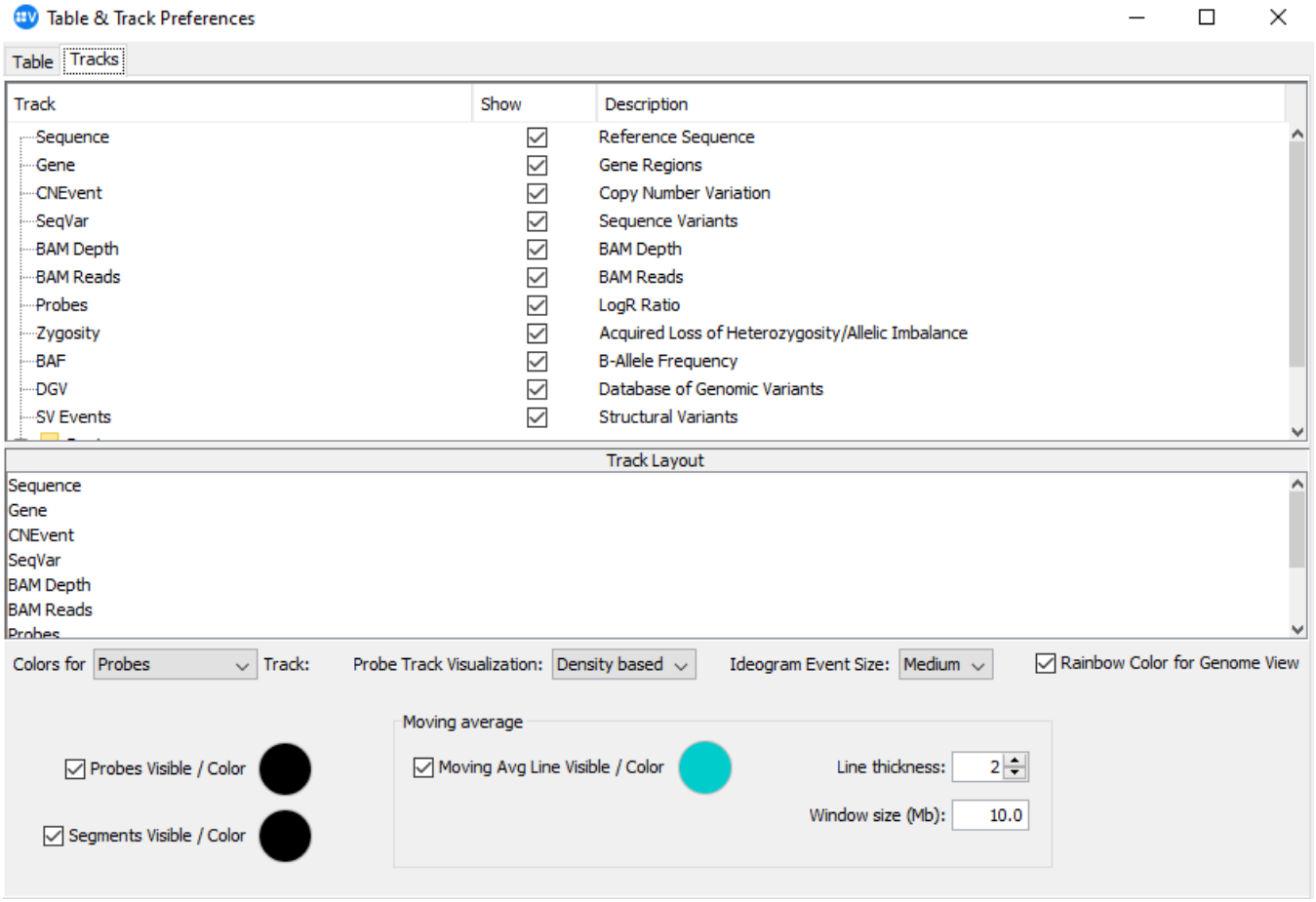


Figure 77. Sample window.

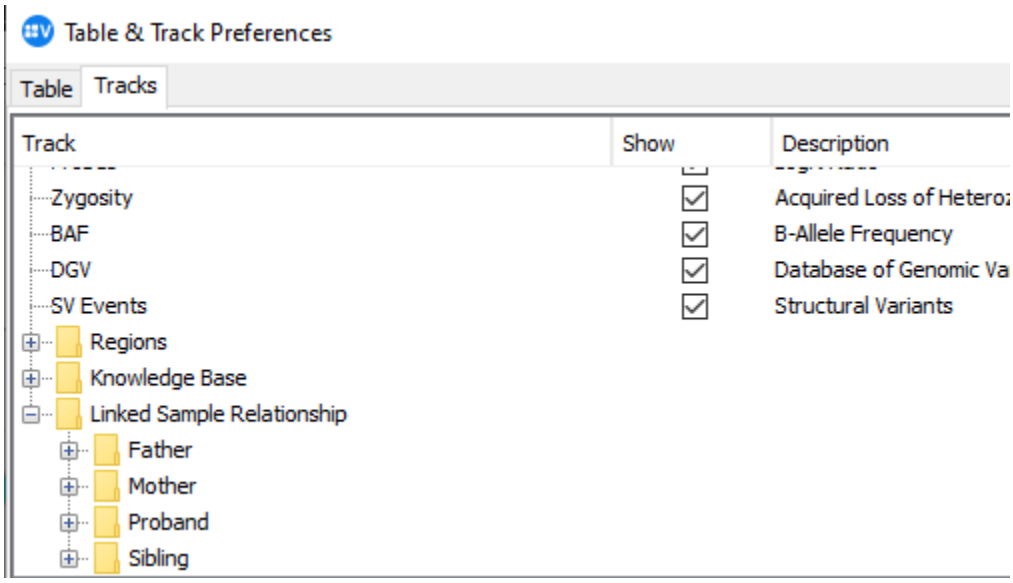


Figure 78. Adjusting columns.

If working with related samples, open the **Linked Sample Relationship** folder to select the option to view data from the additional samples in a single pane. For example, in **Figure 79**, the CN events for the mother, father, and proband will be displayed in the same window as the sample under review.

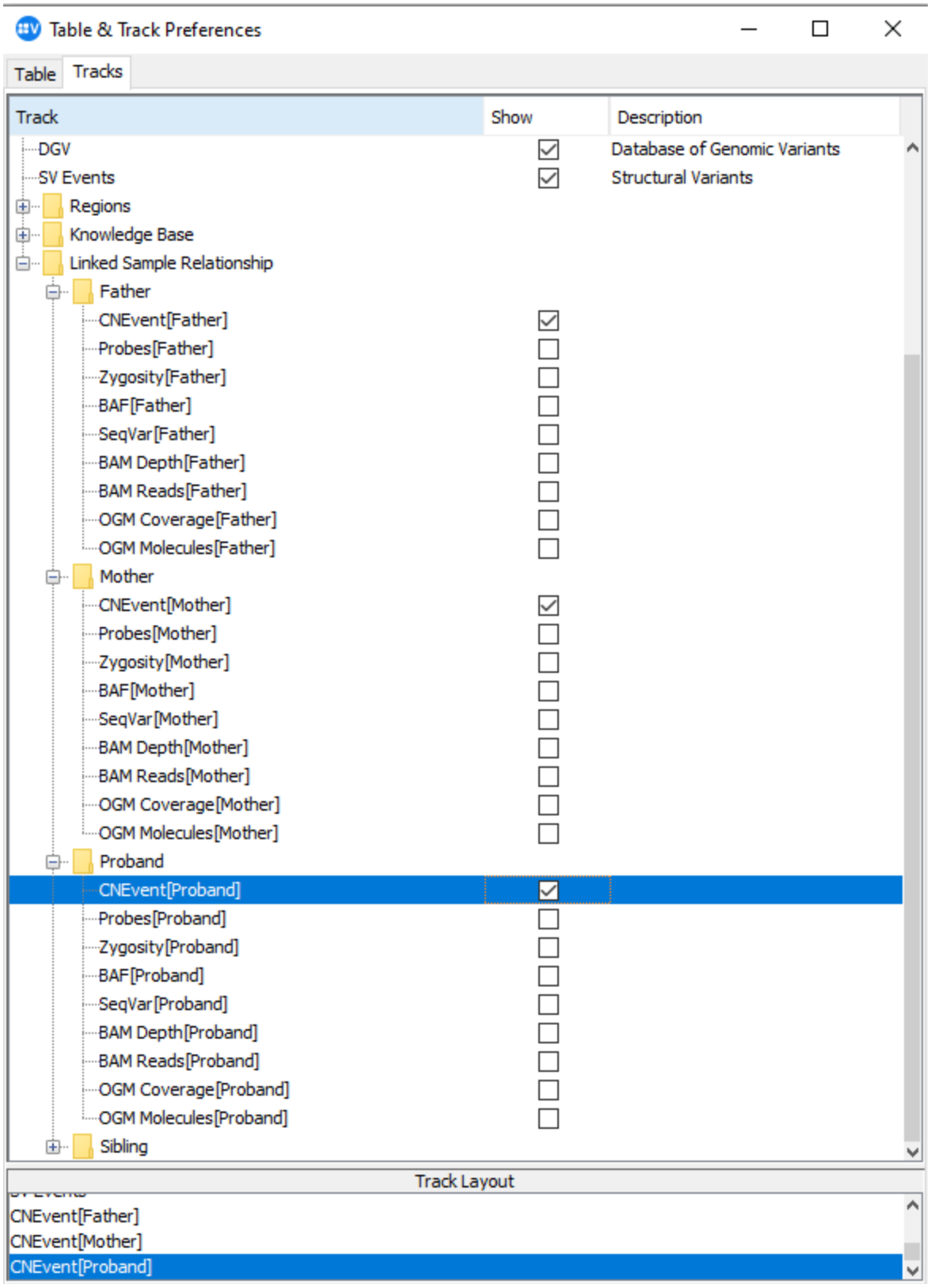


Figure 79. Mother, father and proband CN events.

If the sample has sequence variant and BAM information, then additional tracks will be displayed. **Figure 80** is an example of an NGS sample processed for CNVs (via BAM file) and with sequence variants (via VCF file), therefore additional tracks are available (**BAM Depth**, **BAM Reads**, and **SeqVar**). In **Figure 81**, CN Events for the biological parents of the sample are shown.

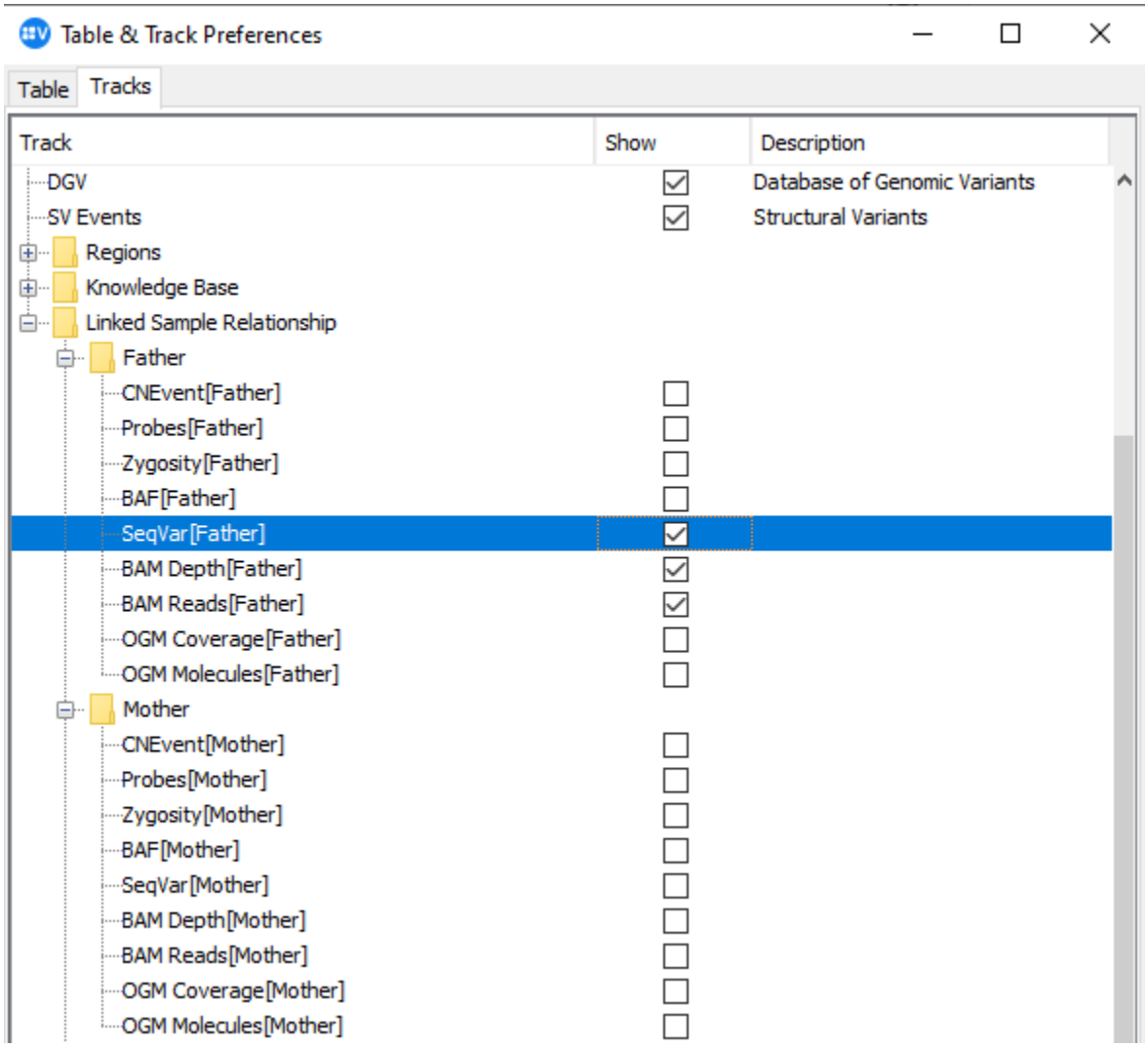


Figure 80. Additional tracks.

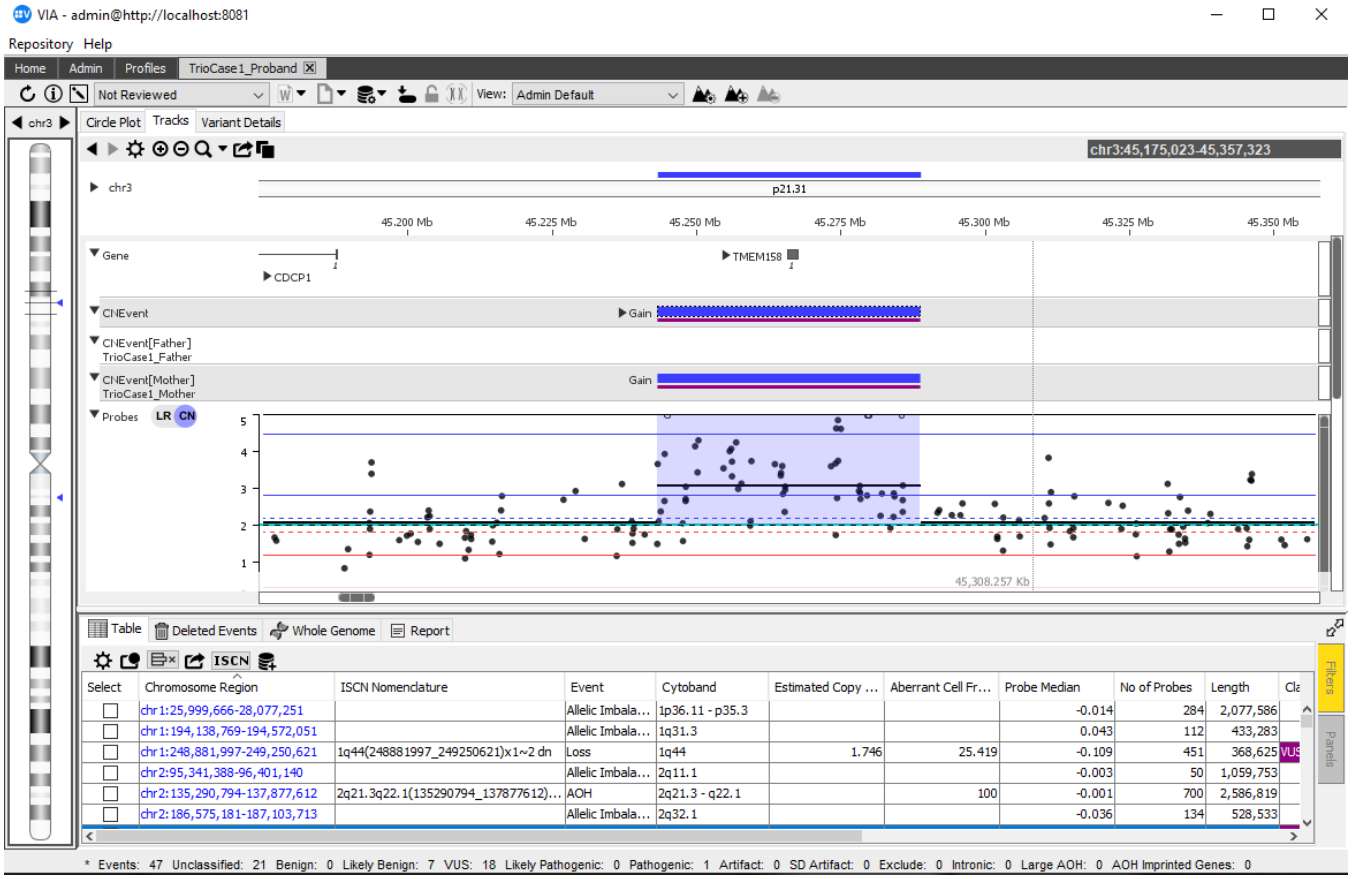


Figure 81. Both the father's and mother's copy number events are displayed.

The order of the tracks is changed in the Track Layout section by highlighting a track name and dragging up or down in the list, as shown in **Figure 82**.

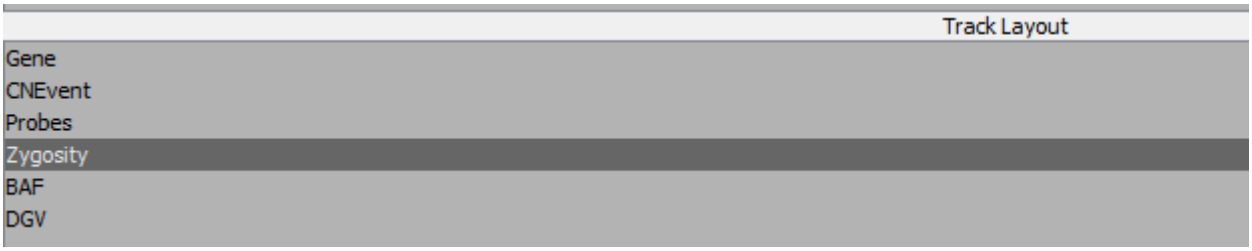


Figure 82. Track layout.

Below the track selections is a section to choose visibility and color representation for the probes and BAF plots as well as the option to display the probes in the Genome View in rainbow colors rather than in gray, as seen in **Figure 83**. To adjust display for the various probes and BAF displays, select a track from the dropdown and then the checkboxes to hide (unchecked) or display (checked) the probes, segments, and the moving average line. Clicking on the colored circles brings up a color chooser where users can change the display color for the respective item. The thickness and window size of the moving average line can be adjusted as well.

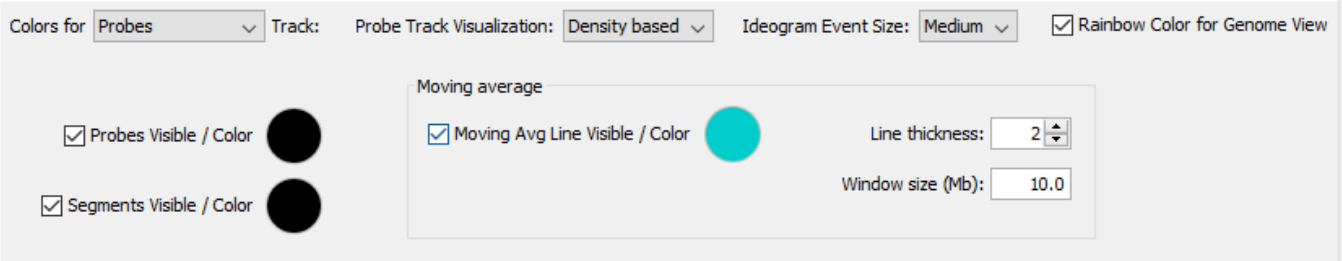


Figure 83. Visibility and color representation.

Visualization in the **Probes** track can be changed to either a Classic or a Density based display, as shown in **Figure 84** and **Figure 85**. Classic displays all probes in the same color. The density-based visualization implements a gray color gradient to depict density of probes in a location. Darker gray indicates many probes in that location and lighter colors indicate fewer probes. This enhances the ability to see density of probes when zoomed out where one pixel could represent one probe or one hundred probes. With a single gradient, there is no indication of the number of probes in that location. With the gradient, it is easy to tell that there are more probes depicted by one pixel in one location versus one pixel in another location.

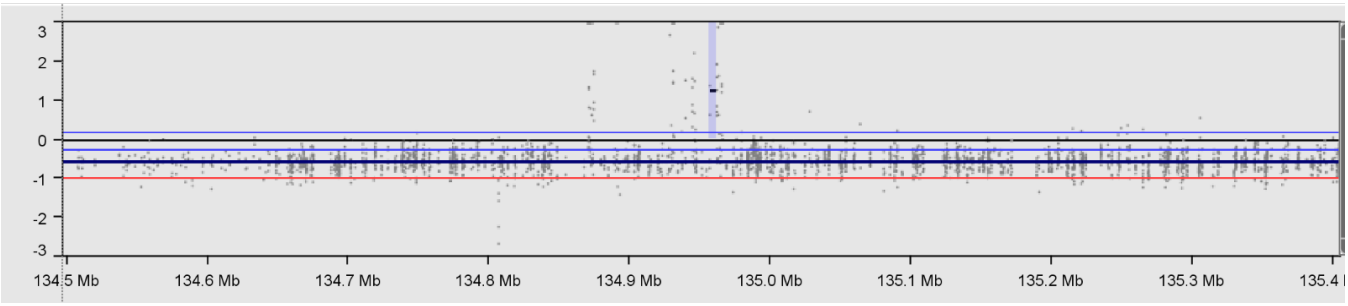


Figure 84. Classic.

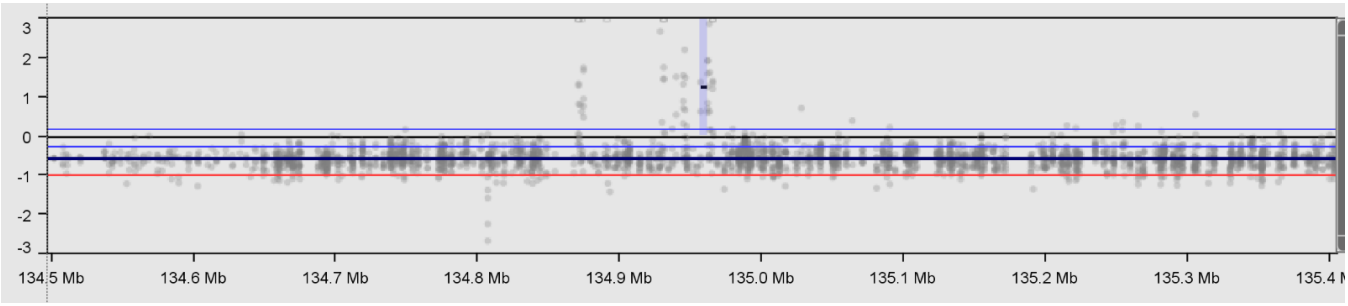


Figure 85. Density-based.

Saving View Preferences

When a sample is first opened, the view layout is the default layout defined by the Administrator. Different types of samples or the review stage can require different views for the most efficient review and interpretation process. There are many customizations for layouts and views for different users, sample types, or even on an individual sample basis. Some preferences require certain user privileges to save them.

COMPONENTS OF VIEWS

View settings affect the following components:

- Table Layout
- Similar Previous Cases Query
- Tracks Layout
- Filter Pipeline
- Display Layout

When creating or altering settings for a view, the user chooses which settings to save, as shown in **Figure 86**. To be prompted to save view preferences before closing a sample, mark the checkbox at the bottom.

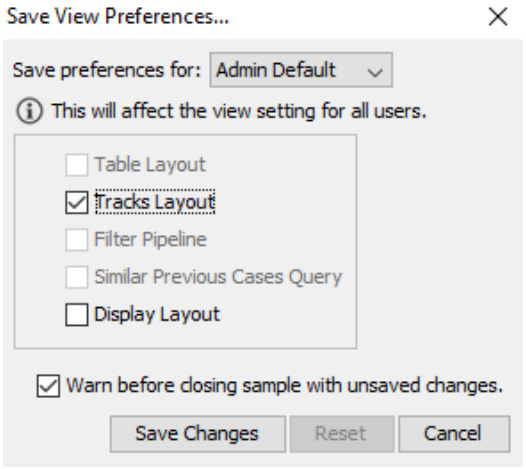


Figure 86. Creating or altering settings for a view.

TYPES OF VIEWS

There are several different types of views available, and each has its own features/functions. They are displayed in the **View** dropdown in the **Sample Review** tool bar, shown in **Figure 87**.

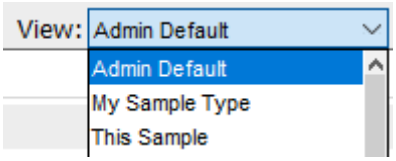


Figure 87. Different views.

The following reserved view types are available by default for each sample. In addition, any number of custom views may be created and saved by a user.

- Admin Default
- My Sample Type
- This Sample

Admin Default: This view is intended to reflect the default view preferences set up by the admin for visualization of all samples within the same sample type.

- Set by the Admin for a specific sample type.
- An Admin can only alter view settings.
- Settings apply for all users.

My Sample Type: This view is intended for a user's own visualization preferences for all samples of the same sample type.

- Any user can alter/save this view.
- Settings apply only for the current user.
- Settings apply for a sample type (sample type of the current sample)

This Sample: This sample view is shared for an individual sample so that each user sees the same sample preferences for a specific sample.

- Requires the following user permission to be enabled for a user to edit/save the **This Sample view** (Ability to save view preferences for a sample).
- The sample must be in **Edit** mode to save this view.
- Settings apply for all users.
- Settings apply only to the current sample.
- A sample must be in this view to lock the sample after final review.

CUSTOM VIEWS

- Can be created by any user.
- Only available to the user that created that custom view; cannot be shared with other users.
- View applies to all sample types in the database.
- Any number of views can be created.
- Appears at the bottom of the list after the other default view types, in order of creation (most recently created displayed first in the list of custom views). See **Figure 88**.
- Can be deleted.

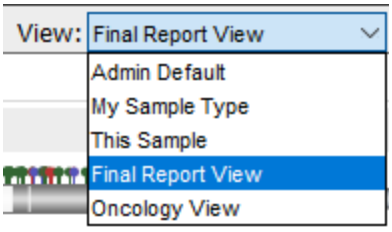


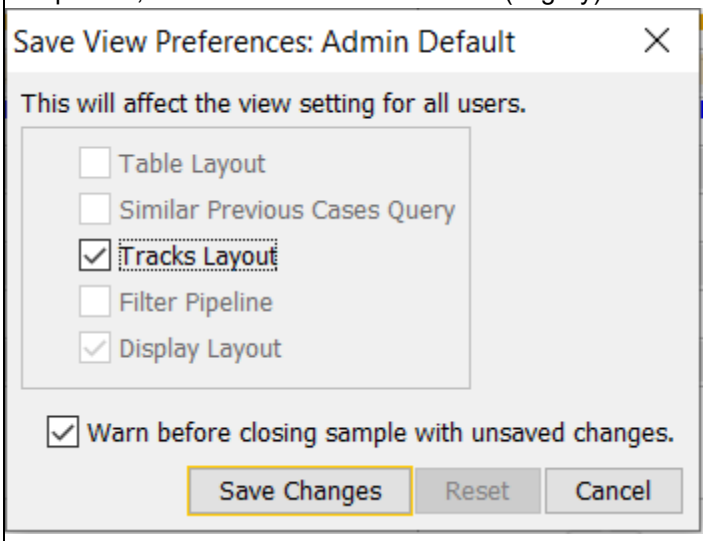
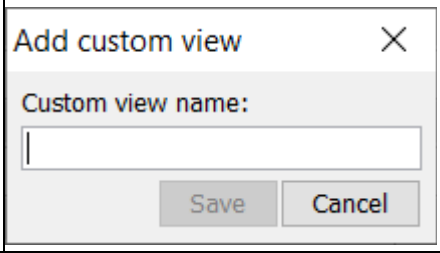
Figure 88. A custom view.

ADDING/SAVING/DELETING VIEWS

- Creation and management of views is accomplished via tools in the **Sample Review** window.

- The tools may be enabled/disabled based on the user privileges and potential requirement of the sample to be in **Edit** mode.

Table 2. Sample Review tools.

Save View Preferences	Opens the **Save View Preferences** dialog for the selected view to choose which settings to save. Mark off checkboxes to save settings for those components. If no changes were made to the view settings for a component, the checkbox will be disabled (in gray). 
Add Custom View	Allows user to create a custom view. 
Delete Custom View	Allows user to delete the custom view currently selected in the dropdown.

Actions for Events

EDITING/ADJUSTING AN EVENT

It may be necessary to change the boundaries of an event call or add or delete a call based on visual inspection of the probes. Only users with editing privileges will be able to manually modify an event.

NOTE: Sequence variant events cannot be modified.

Editing mode is turned on by clicking on the **Edit Sample** button. This locks the sample for the current user so that no other user can simultaneously make changes to the sample. Other users will be able to view the sample at the same time but will not be able to edit it. There are two ways to adjust the boundaries of an event.

1. By expanding the existing event boundaries by dragging the rectangle around the event. First click on the event to display the dashed rectangle around the event (see **Figure 89**). Move the mouse over the boundary in the **CN Event** track until the pointer turns into a double-sided white arrow, then click and drag the boundary to the correct location. Once the boundary line has been moved, a window opens showing coordinates for the current and new region. Users can edit the values in here if needed and then click **Modify** to change the boundary.

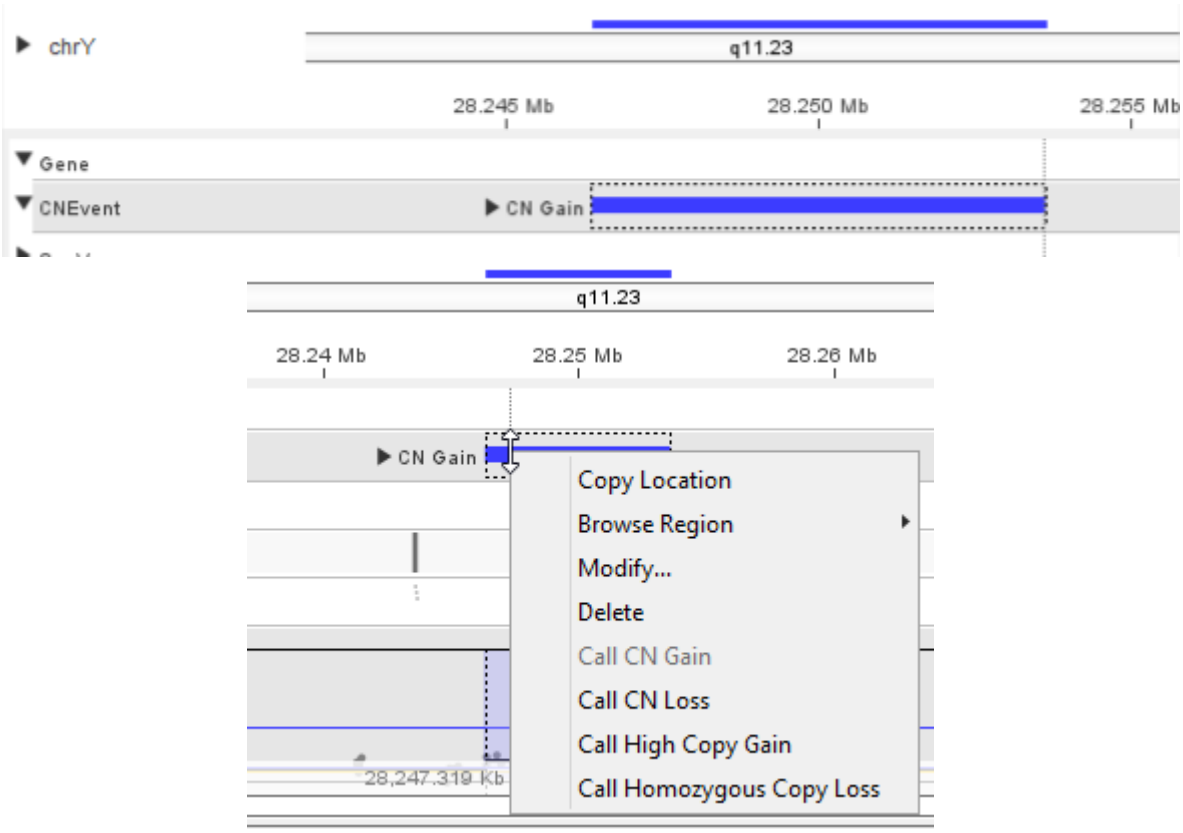


Figure 89. Dashed rectangle around the event.

2. Right click on the event and select. A window displaying the coordinates is launched:

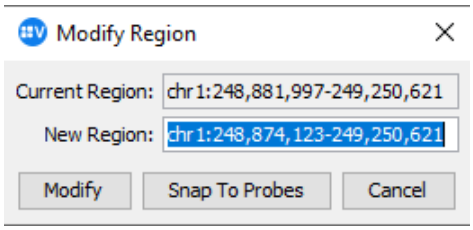


Figure 90. Modify Region window.

Either manually change the coordinates in the **New Region** field and click **Modify** or to automatically select the closest probes, click **Snap to Probes**, which will choose the midpoint of the closest probes to the coordinates in the **Current Region** field. Manual modifications of the event boundaries will be automatically recorded in the **Notes** section for the call, and if the sample type has a decision tree associated with it, the software will ask if auto-classification should be run again on this adjusted event. A notification will pop up (seen in **Figure 91**).

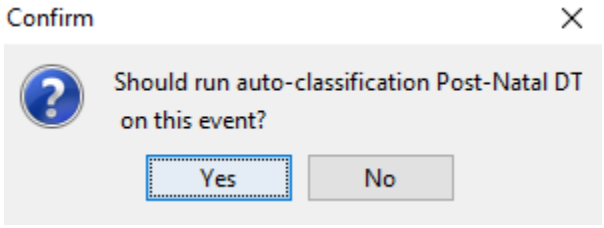


Figure 91. Auto-classification notification.

If **Yes** is selected, another notification will show that the decision tree is running:

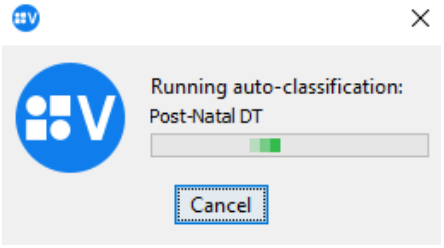


Figure 92. Decision Tree running.

NOTE: There will still be an opportunity to cancel the operation while the decision tree is running.

DELETING AN EVENT

NOTE: For events to be deleted the sample must be in edit mode. Not all users may have been given permission for this function to be enabled. The sequence variant events cannot be deleted.

Click on the **Edit** button to start editing the sample. To select an event for deletion via the browser, right-click on an event to bring up the context menu, as in **Figure 93**, and select **Delete**. A prompt will appear to confirm deletion.

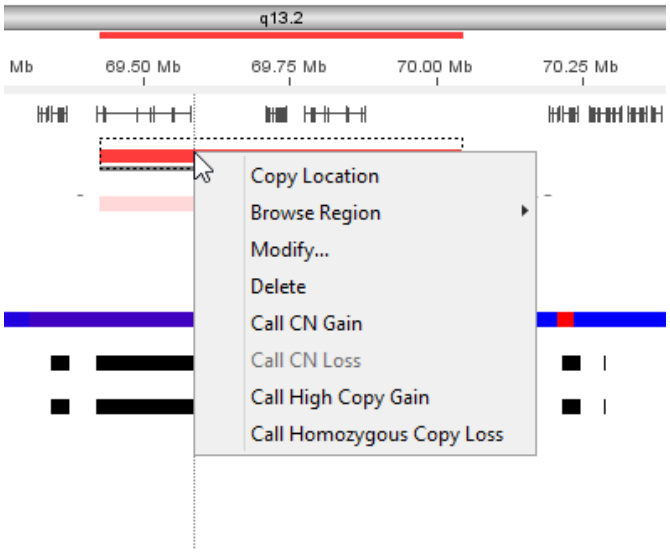


Figure 93. The context menu.

Deleting an event can also be accomplished via the table by clicking on the appropriate row in the table and then clicking the **Delete Events** button. If an event is deleted, another user can see what was deleted by clicking on the button in the table. The data table will then display a list of the deleted events and the icon will change to orange, as seen in **Figure 94**. The **Notes** column will show who deleted the event and when.

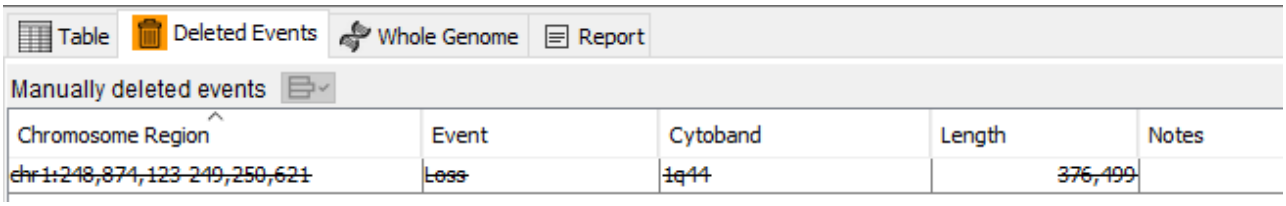


Figure 94. Orange **Deleted Events** icon.

RECOVERING DELETED EVENTS

NOTE: For events to be restored the sample must be in edit mode. Not all users may have been given permission for this function to be enabled.

From the manually deleted events view, select rows for the events to be restored and click on the **Restore** button. These events will be restored and a notation of when and by whom the event was restored will be made in the **Audit log** column of each event.

ADDING AN EVENT

NOTE: For events to be added the sample must be in edit mode. Not all users may have been given permission for this function to be enabled.

Adding a CNV or allelic event call: To add a CNV or allelic event call, zoom in to the region desired. Then choose the **Selection** tool, as seen in **Figure 95**, from the **Tools** menu.

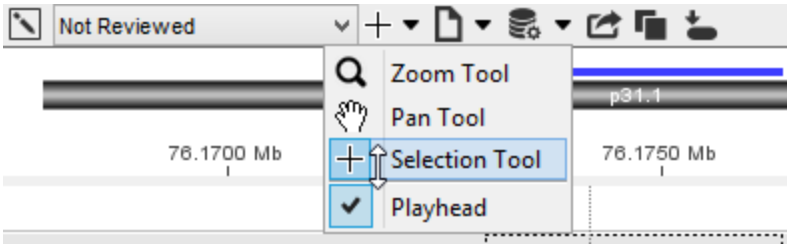


Figure 95. The **Selection** tool.

Select the region in the **Probes** track (for copy number calls) or BAF track (for allelic events) where the call should be added. Once the correct region is outlined, right click, and select the call to be added. Users can further adjust the boundaries as outlined in the section editing/adjusting an event above.

Figure 96 displays an editing context menu for adding/deleting copy number calls. **Figure 97** displays the editing menu for the same functions but for allelic event calls.

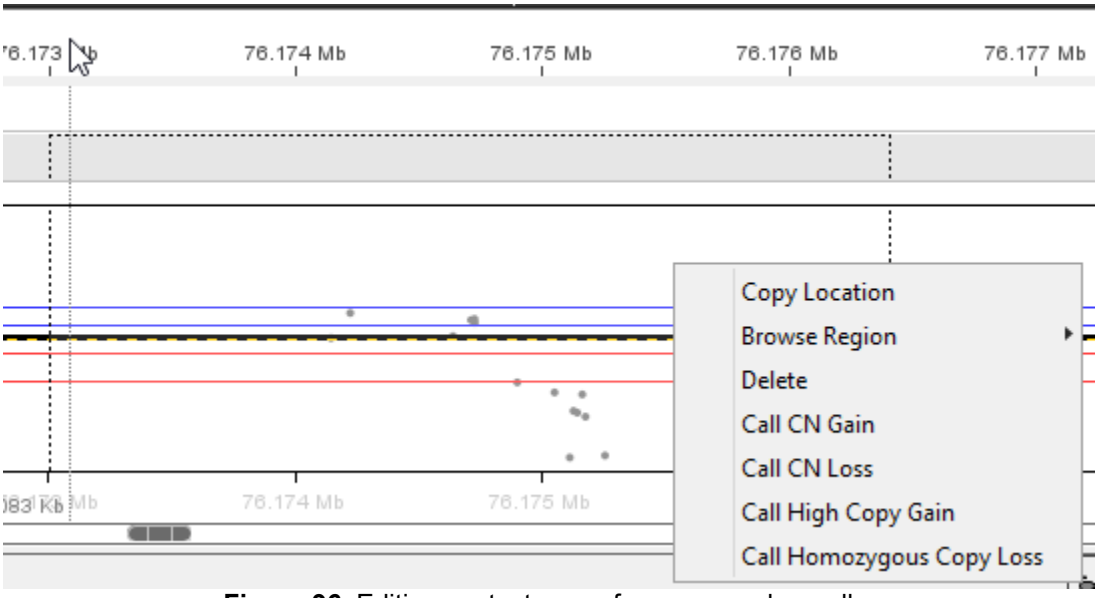


Figure 96. Editing context menu for copy number calls.

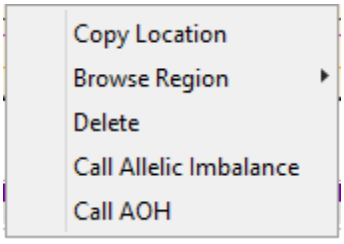


Figure 97. Editing menu for adding/deleting allelic event calls.

A window with event coordinates is launched:

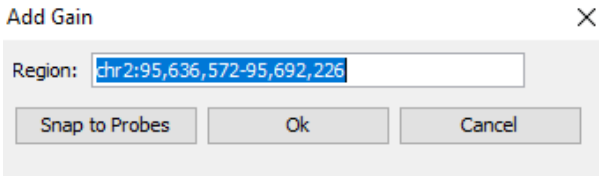


Figure 98. Event coordinates.

Click **OK** to add the call with the coordinates displayed in the **Region** field. Or click **Snap to Probes** to automatically select the closest probes for the call boundaries. Clicking **Snap to Probes** will choose the midpoint of the closest probes to the coordinates displayed in the Region field.

Once a call is selected, and if there are any decision trees associated with the sample type, a prompt for running the pre-classification decision tree on this event will appear, as in **Figure 99**.

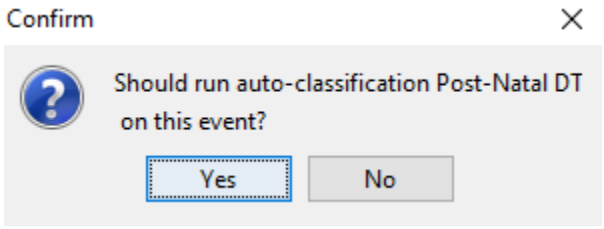


Figure 99. Auto-classification prompt.

After selecting **Yes**, another prompt will show progress of the pre-classification and there is an option to cancel at this point.

ADDING A SEQVAR EVENT

To add an event (must be in edit mode), select the button from the tools bar in the table. A window opens, as seen in **Figure 100**, where details about the variant can be entered (no commas in start/end field). Click **Add** to add the sequence variant. Manual calling of events will be automatically recorded in the **Audit Log** section for the event.

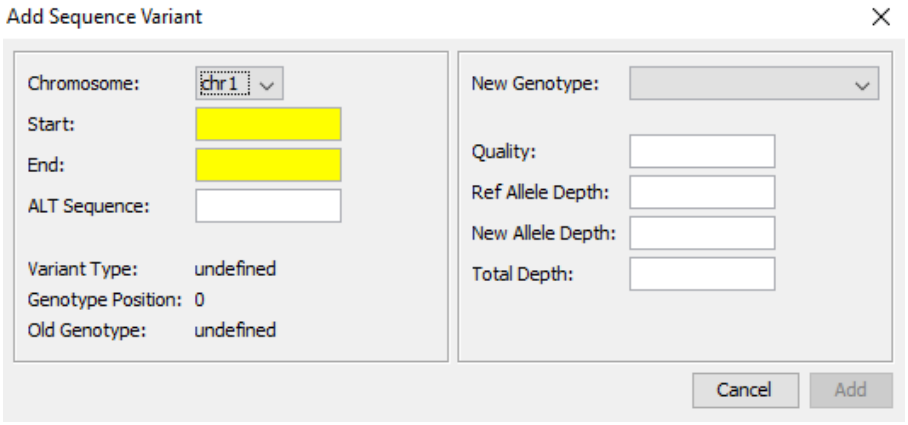


Figure 100. Window for variant details.

Results Table Navigation and Toolbar

The lower panel of the window has tabs displaying the table and other information. At the top within each tab is a row of tools applicable to the contents of that tab, as seen in **Figure 101**.

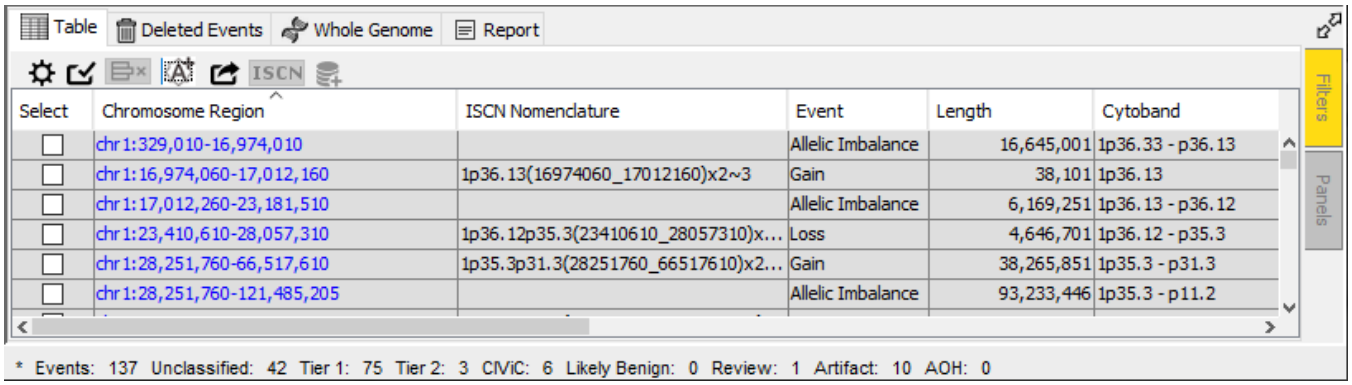


Figure 101. Results table navigation and **Toolbar**.

TABLE VIEW

The **Table** view is selected by clicking on the **Table** tab. At the top are the tools available for this tab, shown in **Figure 102**.

Figure 102. Tools available for the **Table** tab.

The data table contains information about each event including, but not limited to, length, location, classification, and notes. The table column content is customizable.

Chromosome Region	Event	Genes	Pop. Allele ...	Transcripts	Total Depth	Variant Fre...	Quality
chr8:7,213,638-7,779,382	CN Gain	ZNF705G, DEFB4B, DEFB10...					
chr17:44,175,785-44,302...	CN Gain	KANSL1, KANSL1-AS1					
chr9:847,069	SNV	DMRT1	0.51 (1KG SAS)	missense_variant	75	45.333	25
chr9:13,188,785	SNV	MPDZ	0.82 (1KG SAS)	probably_damaging	69	56.522	25
chr9:34,635,598	SNV	SIGMAR1		NMD_transcript_variant	51	100	10
chr9:35,705,570	SNV	TLN1	0.017 (ExAC ...	missense_variant	80	46.25	25
chr9:71,661,376	SNV	FXN	0.2 (1KG SAS)	missense_variant	85	51.765	25
chr9:90,260,847	SNV	DAPK1		probably_damaging	76	46.053	25
chr9:97,863,965	SNV	FANCC	0.055 (ExAC ...	regulatory_region_variant	29	65.517	25
chr9:131,859,552	SNV	CRAT		missense_variant	84	13.095	17
chr9:137,779,306	SNV	FCN2		3_prime_UTR_variant	134	100	10
chr9:138,669,210	SNV	KCNT1	0.8 (1KG EUR)	synonymous_variant	182	100	10

Figure 103. Table View data table.

To sort samples, click on the column header. Clicking again will sort in reverse. To sort on multiple columns, hold down the **CTRL** key while clicking on the header of different columns. The size of the arrows (larger to smaller) indicates which column is the primary, secondary, and so on, with respect to sorting. Clicking on an event in the **Table** view will zoom in on that event in the graphical display (in the tracks).

SEQVAR ANNOTATIONS

For samples of the NGS class, the table in **Figure 104** can have many columns such as individual transcript annotation columns, a transcript overview column (displaying the most severe consequence), and a transcript ID column. These annotation columns will be filled automatically. If the gene has a single canonical transcript either in RefSeq or Ensembl, this will be the selected transcript. If both databases have a single canonical transcript, then RefSeq will be used by default. If there are multiple canonical transcripts, the most damaging one is selected; if all consequences are of the same level of severity, the longest transcript is selected. Once a transcript is selected, all resulting annotation details are based on the selected transcript.

- **Transcript Overview:** Displays the most interesting (most severe) consequence of the selected transcript. And if that transcript is the canonical one, the text in the column will be displayed in bold. Clicking on the cell opens the **Transcript Information** window.
- **Consequence:** Lists all consequence values for the transcript.

Transcript Overview	Consequence	Transcript ID
Missense variant	Missense variant	NM_199244.2
Stop gained	Stop gained	NM_199244.2
Missense variant	Missense variant	NM_207355.2
Missense variant	Missense variant	NM_207355.2
Frameshift variant	Frameshift variant	NM_207421.3
Inframe deletion	Inframe deletion	NM_207446.2
Missense variant	Missense variant, Splice region variant	XM_001726942.4
Inframe deletion	Inframe deletion	XM_003959926.1

Figure 104. NGS class samples.

There are four interest levels, listed from least interesting (least severe) to most interesting (most severe):

- Not interesting - blue
- Maybe interesting - green
- Interesting - gold
- Very interesting – red

If the transcript has more than one consequence and they belong to different interest levels, then the color used for the transcript ID will be based on the highest interest level represented among the different consequences.

In **Figure 105**, a transcript has two consequences, one that belongs to Not Interesting (blue) and the other to Very Interesting (red). Since Very Interesting is the more severe consequence, the transcript name (ENST00000437966) is displayed in red. If the sequence variant file includes protein predictions (e.g., via PolyPhen, SIFT), then those will be color-coded as well.

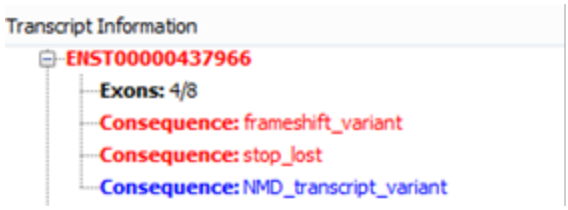


Figure 105. Transcript with two consequences.

TRANSCRIPT INFORMATION WINDOW

The **Transcript** window, shown in **Figure 106**, lists all transcripts from RefSeq and Ensembl along with annotations for each transcript. Clicking on the transcript name expands the panel and displays additional information about the transcript including the variant consequences. The consequences are color-coded to represent the interest level (severity) of the consequence.

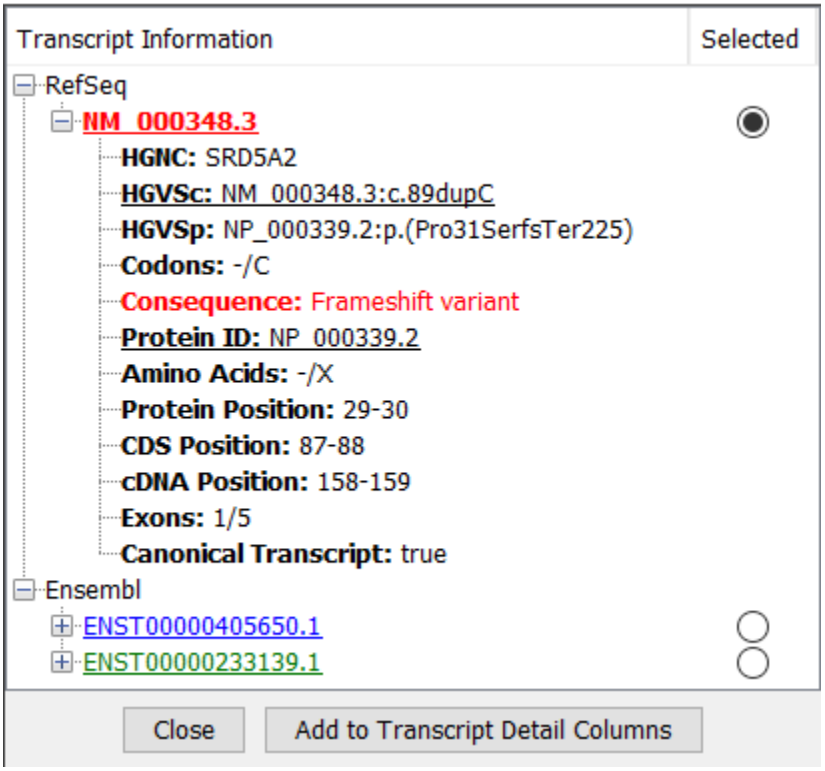


Figure 106. Transcript window.

The **Selected** radio button shows which transcript is being used for annotations. Various transcript annotation columns of the table display values associated with the selected transcript. A default transcript is selected automatically (described earlier) after the sample is processed.

To change the selected transcript in the **Transcript Information** window, the Sample must be in **Edit** mode. If the user is not in **Edit** mode, the buttons will be inactive. Once a new transcript has been selected, click **Add to Transcript Details Columns** to save the new selection. An entry will be made in the **Audit Log** indicating that

the selected transcript has been changed. All **Transcript Details** columns will now display annotations associated with the new selected transcript, as shown in **Figure 107**.

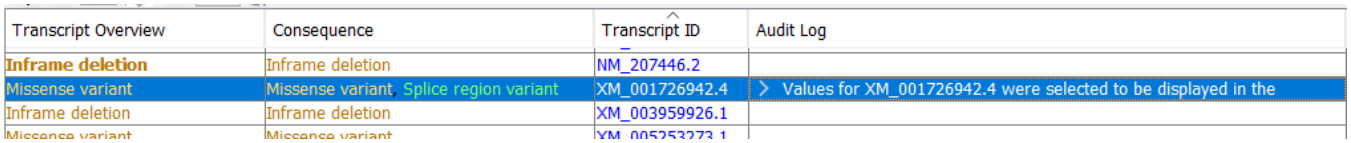


Figure 107. Transcript details.

REGULATORY REGIONS

The table in **Figure 108** has a column entitled **Regulatory Feature** which displays the consequence for regulatory regions (if these are available in the VCF or JSON file). If there is regulatory feature information, a blue hyperlinked regulatory_region_variant text will be displayed and clicking on it will open a new window with details on the regulatory region features. Marking the checkbox in **Edit** mode will allow addition of the annotation to the **Notes** field.

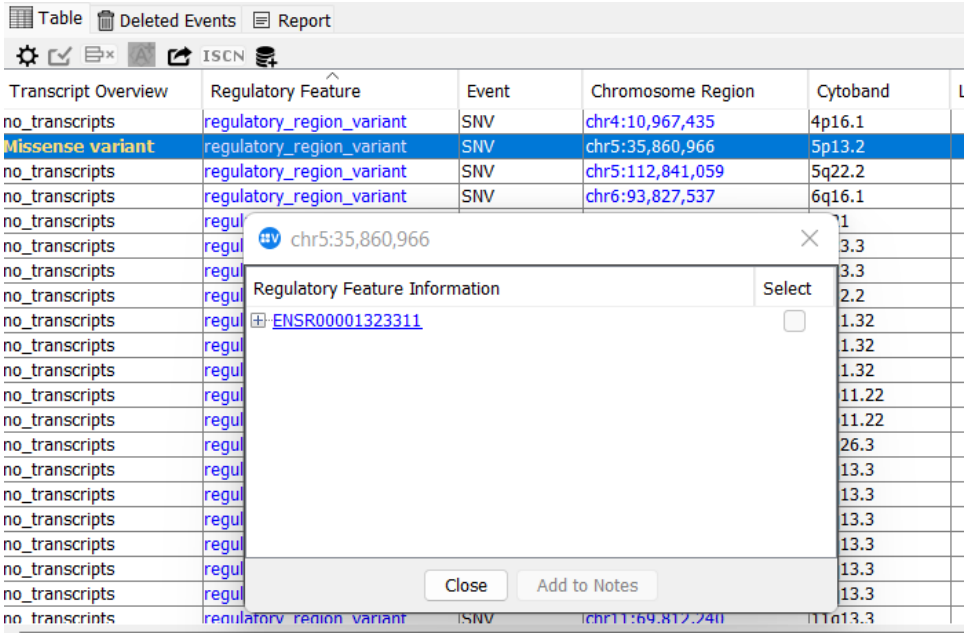


Figure 108. Regulatory region details.

POPULATION ALLELE FREQUENCY

This column displays the highest population allele frequency, population, and the source. It is hyperlinked and clicking the link opens a new window providing population allele frequencies from several data sources. Additional information showing allele counts and number of homozygotes is also included where available, shown in **Figure 109**.

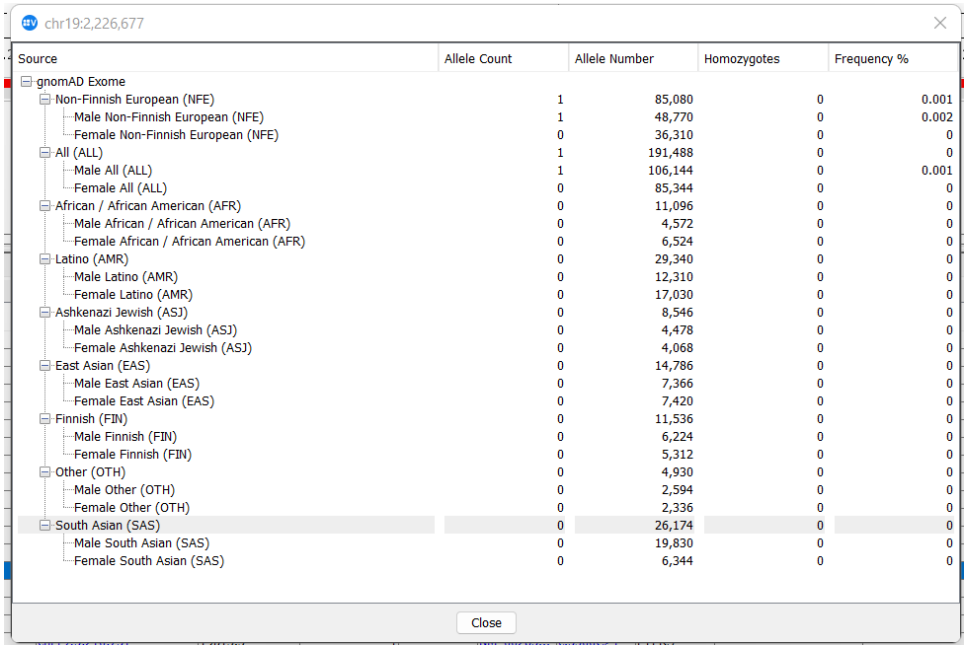
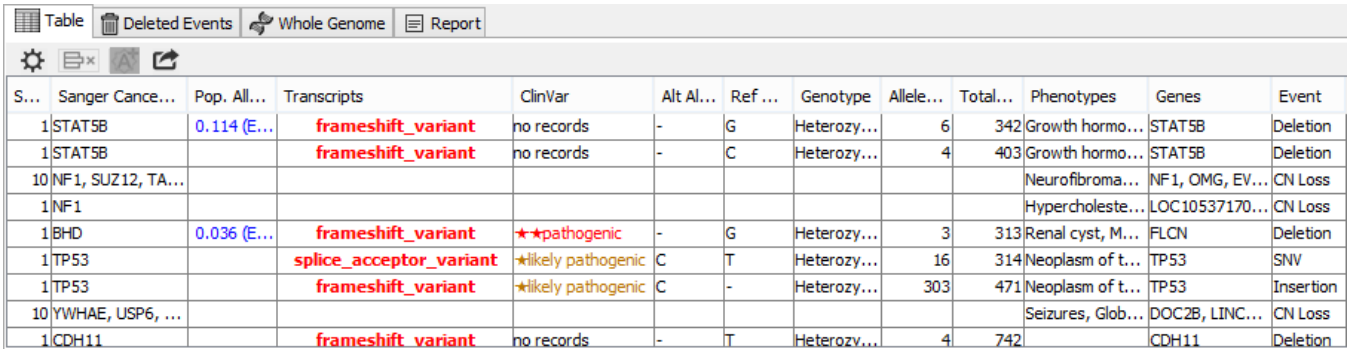


Figure 109. Population allele frequency.

In the **ClinVar** column, if the input sequence variant file contains this kind of information, then the column in the table will show the annotations. If no records are available, the notation states no records. Both scenarios are shown in **Figure 110**. If information is available, the classification is displayed along with the star rating for review status and the text is hyperlinked to a window containing **ClinVar** details.



S...	Sanger Cance...	Pop. All...	Transcripts	ClinVar	Alt Al...	Ref ...	Genotype	Allele...	Total...	Phenotypes	Genes	Event
1	STAT5B	0.114 (E...	frameshift_variant	no records	-	G	Heterozy...	6	342	Growth hormo...	STAT5B	Deletion
1	STAT5B		frameshift_variant	no records	-	C	Heterozy...	4	403	Growth hormo...	STAT5B	Deletion
10	NF1, SUZ12, TA...									Neurofibroma...	NF1, OMG, EV...	CN Loss
1	NF1									Hypercholeste...	LOC10537170...	CN Loss
1	BHD	0.036 (E...	frameshift_variant	★★pathogenic	-	G	Heterozy...	3	313	Renal cyst, M...	FLCN	Deletion
1	TP53		splice_acceptor_variant	★likely pathogenic	C	T	Heterozy...	16	314	Neoplasm of t...	TP53	SNV
1	TP53		frameshift_variant	★likely pathogenic	C	-	Heterozy...	303	471	Neoplasm of t...	TP53	Insertion
10	YWHAE, USP6, ...									Seizures, Glob...	DOC2B, LINC...	CN Loss
1	CDH11		frameshift_variant	no records	-	T	Heterozy...	4	742		CDH11	Deletion

Figure 110. The ClinVar column.

The classification terms used are those recommended by ACMG guidelines and are also color coded to indicate significance. Classification terms and color coding in order of significance (most to least):

- pathogenic - red
- likely pathogenic - gold
- uncertain significance - green
- likely benign - blue
- benign – blue

Stars next to the classification terms indicate the review status (star rating used by ClinVar), shown in **Table 3**.

Table 3: ClinVar Star Ratings indicating review status.

Number of stars	Description and review statuses
none	No submitter provided an interpretation with assertion criteria (no assertion criteria provided), or no interpretation was provided (no assertion provided) .
one	One submitter provided an interpretation with assertion criteria (criteria provided, single submitter) or multiple submitters provided assertion criteria but there are conflicting interpretations in which case the independent values are enumerated for clinical significance (criteria provided, conflicting interpretations).
two	Two or more submitters providing assertion criteria provided the same interpretation (criteria provided, multiple submitters, no conflicts) .
three	reviewed by expert panel
four	practice guideline

In cases where the variant has been classified with more than one term, the classification with the highest review status is listed first; the most significant classification will also be displayed in the cell but second to the one with the higher review status.

The benign classification has a higher review status (two stars = criteria provided, multiple submitters, no conflicts) and is listed before uncertain significance which is of higher significance but with a lower review status (one star = criteria provided, single submitter). ★★benign (★uncertain significance)

Sorting on the **ClinVar** column is based on the classification significance rather than star rating, as seen in **Figure 111**. An entry with three stars (likely benign) and one star (pathogenic) will be sorted based on the classification significance such that the entry will be sorted together with other pathogenic variants. This is so that a variant is brought to the reviewer's attention and can be seen even if only a single submitter classified the variant as pathogenic.

S...	Sanger Cance...	Pop. All...	Transcripts	ClinVar	Alt Al...	Ref ...	Genotype
1	TP53		splice_acceptor_	★likely pathogenic	C	T	Heterozy...
1	TP53		frameshift_varia	★likely pathogenic	C	-	Heterozy...
10	YWHAE, USP6, ...						
1	CDH11		frameshift_varia	no records	-	T	Heterozy...
7	HERPUD1, CDH...						
1	CYLD						
1	CYLD						
1	MYH11	0.868 (E...	frameshift_variant	★★benign (★uncertain significance)	G	-	Heterozy...
1	CREBBP						

Figure 111. ClinVar column displaying one variant with multiple classification terms.

Clicking on the **ClinVar** cell will open a new window with further details from **ClinVar**. Clicking on likely pathogenic, shown in the first row of **Figure 111**,opens the window shown in **Figure 112** below. Each record is hyperlinked to its page on the **ClinVar** website. Marking the checkbox column and clicking **Add to Notes** will add the information to the **Notes** column of the **Results** table.

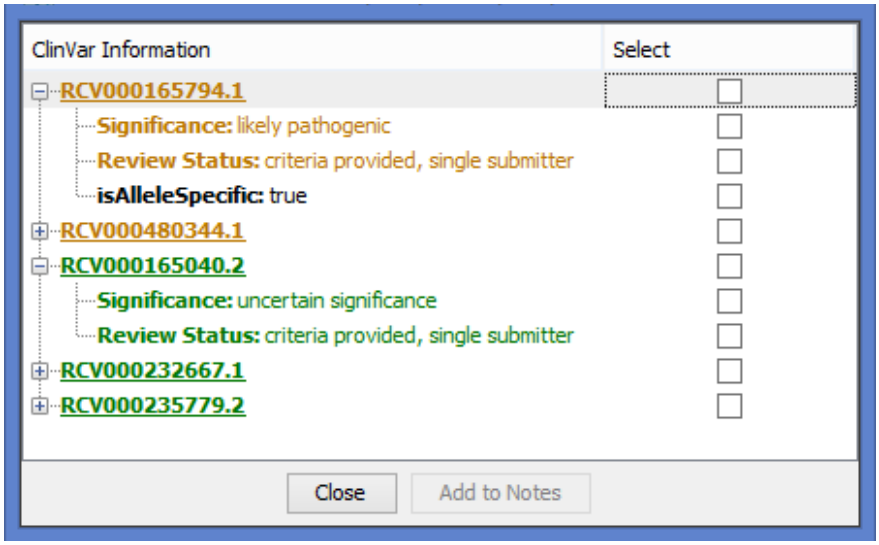


Figure 112. Further ClinVar details window.

THE COSMIC COLUMN

Most sequence variants files will not have this information. Old Nirvana annotations had COSMIC information but the latest does not. For the COSMIC column (data obtained from sequence variant files that have COSMIC annotations), the value in the cell (in **Table 4** and in the pop-up window) indicates that the COSMIC field isAlleleSpecific is true.

Table 4. COSMIC information.

PAF	COSMIC	Genes	Transcripts	ClinVar	Event	Chromoso...
	0 records	LOC10272...	frameshift_v	no records	Insertion	chr1:17,086...
	0 records	MST1L	frameshift_v	no records	Deletion	chr1:17,087...
	0 records	LOC10099...	frameshift_v	no records	Deletion	chr1:144,91...
	3 records	LOC10099...	frameshift_v...	no records	Deletion	chr1:145,01...
	1 samples	LOC10192...	frameshift_v	no records	Deletion	chr1:245,13...
	3 samples	LOC10192...	frameshift_v	no records	Insertion	chr1:245,13...
	0 records	ZNF806	frameshift_v...	no records	Deletion	chr2:133,07...
	0 records	ZNF806	frameshift_v...	no records	Insertion	chr2:133,07...

Clicking on the hyperlink opens a new window with further details on records or samples based on what is present in the **COSMIC** column. Marking off the checkbox column will add the information to the **Notes** column of the results table.

In silico prediction columns: If the sequence variants file loaded has *in silico* predictions, these will be displayed in individual columns in the table. If the columns are not visible, use the table preferences to unhide these columns. Some fields may be empty, and this could be due to the transcript selected. The annotator used for annotating a VCF file also impacts which fields will have data. For example, if a JSON file is loaded (annotated with Nirvana outside of the VIA pipeline), it will likely only have PolyPhen and SIFT predictions. VCF files annotated within the VIA pipeline (using the linked Nirvana Annotator) will use dbNSFP for functional predictions so they will have

results from the following additional predictors supported by VIA: FATHMM, MetaLR, Mutation Assessor, MetaSVM, see in **Figure 113**.

FATHMM Prediction	FATHMM Score	MetaLR	MetaSVM	Mutation Assessor	SIFT Prediction	SIFT Score	PolyPhen Prediction	PolyPhen Score	Transcript Overview	Event
									Splice donor variant	Deletion
									Splice region variant	SNV
		Tolerated(0.152)	Tolerated(-0.904)	Medium(2.360)	Deleterious(NM_000553.4)	0.01	Benign(NM_000553.4)	0.275	Missense variant	SNV
									Synonymous variant	SNV

Figure 113. Additional *in-silico* predictors.

MetaSVM, MetaLR, and MutationAssessor predictors do not map to transcripts, so these columns do not have a transcript ID coupled to the consequence value. For these, the score and consequence are in one column. For the other predictors, the scores are provided in a separate column from the consequence + transcript ID. For the others, columns may be empty if the selected transcript did not have a prediction associated with it, shown in **Figure 114**. The highlighted column below shows a RefSeq transcript ID and has no FATHMM Prediction values.

Transcript ID	FATHMM Prediction	FATHMM Score	MetaLR	MetaSVM	Mutation Assessor	SIFT Prediction	SIFT Score	PolyPhen Prediction	PolyPhen Score	Transcript Overview	Event
NM_033084.4										Splice donor variant	Deletion
NM_033084.4										Splice region variant	SNV
NM_000553.4			Tolerated(0.152)	Tolerated(-0.904)	Medium(2.360)	Deleterious(NM_000553.4)	0.01	Benign(NM_000553.4)	0.275	Missense variant	SNV
NM_000548.4										Synonymous variant	SNV

Figure 114. Selected transcript has no associated FATHMM prediction resulting in empty FATHMM Prediction columns.

After selecting an Ensembl transcript, the **FATHMM** columns now have values. In this case, FATHMM scores were only available for some transcripts and if one of these transcripts is not the one selected to be displayed in the table, the prediction and score columns will be empty. The prediction values are color-coded based on the severity of the prediction.

Transcript ID	FATHMM Prediction	FATHMM Score	MetaLR	MetaSVM	Mutation Assessor	SIFT Prediction	SIFT Score	PolyPhen Prediction	PolyPhen Score	Transcript Overview	Event
NM_033084.4										Splice donor variant	Deletion
NM_033084.4										Splice region variant	SNV
ENST00000298139.5	Tolerated(ENST00000	0.59	Tolerated(0.152)	Tolerated(-0.904)	Medium(2.360)	Deleterious(ENST0000029813	0.01	Benign(ENST0000029813	0.275	Missense variant	SNV
NM_000548.4										Synonymous variant	SNV

Figure 115. An example of an Ensembl transcript with FATHMM scores.

The **Transcript Overview** window, shown in **Figure 116**, is opened by clicking on the **Transcript Overview** field and will also display *in silico* predictions and scores where available. For more information on these predictors, please visit the dbNSFP website.

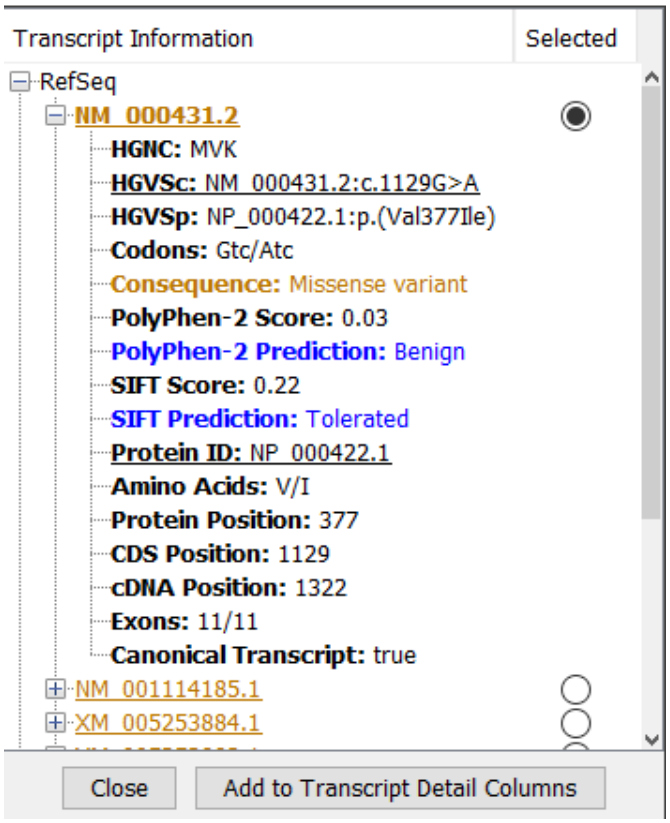


Figure 116. Transcript Overview window.

TABLE LAYOUT

The table layout and columns displayed can be changed via the **Table Preferences** button just above the data table or via the same symbol above the ideogram at the top of the window. Make sure the **Table** tab is selected.

Columns are grouped together into folders for better management and easier selection of columns to display/hide **CN** and **Allelic Events**, **Sequence Variant Events**, **Regions**, and the **Decision Tree**. The **Sequence Variant Events** folder has a **Transcripts** folder for transcript annotations and an ***In-Silico* Predictions** folder for *in silico* predictors.

One can select which columns to display and in what order. Drag headers in the **Column Layout** section at the bottom (see **Figure 117**) to move columns around to re-arrange information. Column widths can be re-sized by dragging the column header edges (cursor will turn into a double headed arrow when moving the mouse over the column header edges).

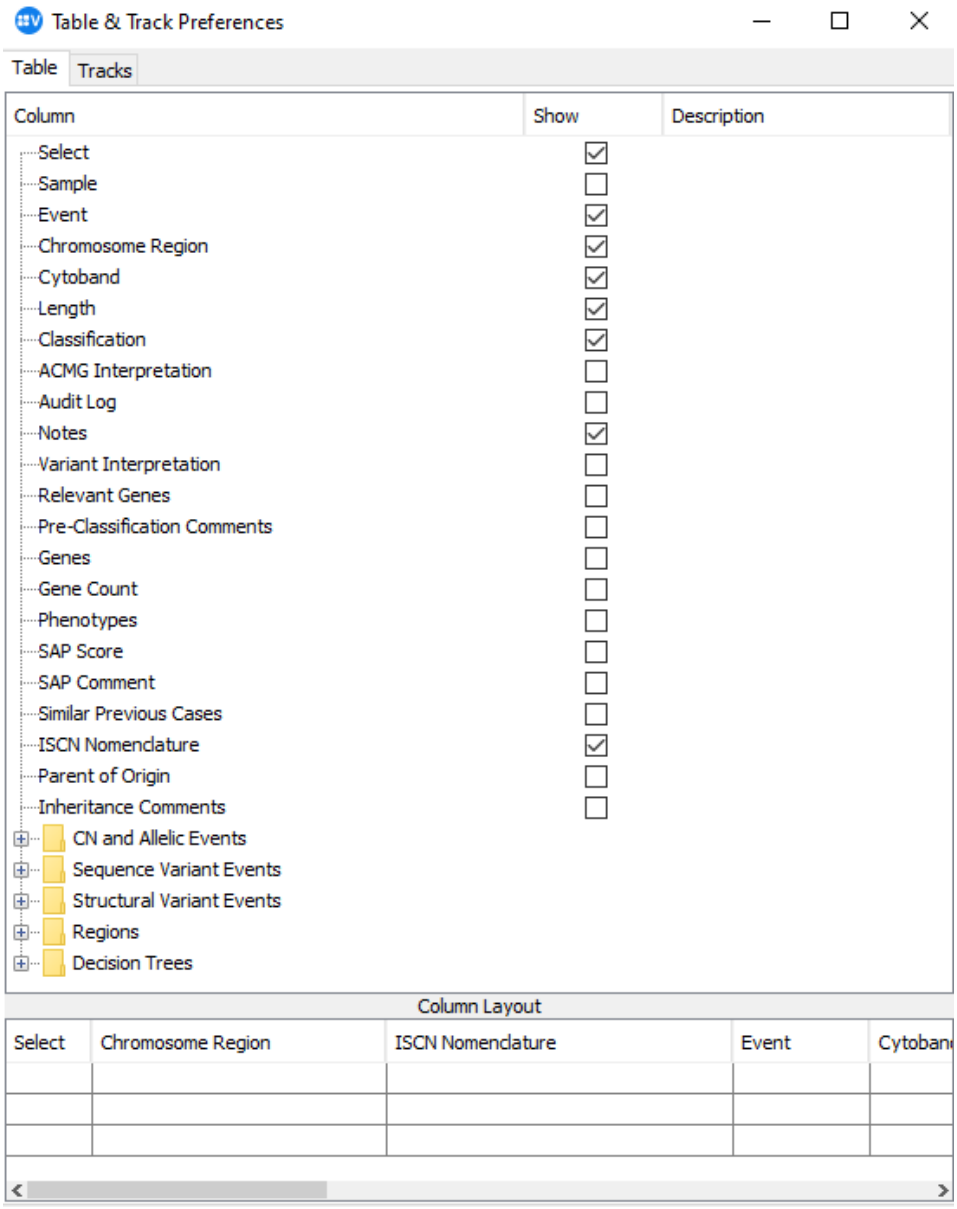


Figure 117. Re-organizing information.

Sample Review Table Columns

Table 5 displays a limited number of columns by default. The user can go into the table preferences to select which columns to display and order the columns. The table below shows the possible columns available in a table. Based on the sample type, only a selection of the columns below may be available. **Table 5** displays the general data for each sample type (available for every sample type).

Table 6 through **Table 10** are for each individual group of results (these will be available depending on the sample type).

Table 5. General sample type data.

Column	Description
Select	Allows user to mark events to export for reporting.
Sample	Name given to the sample in the Sample Descriptor file.
Event	Possible values are **CN Gain**, **CN Loss**, **High Copy Gain**, **Homozygous Loss**, A**OH**, and **Allelic Imbalance** for CN and allelic events and SNV, insertion, deletion, MNV, and indel for SeqVar.
Chromosome Region	Chromosome along with a base pair region in the following format: chr8:172,199-300,002.
Cytoband	The cytoband this region covers.
Length	Length of the region.
Classification	The classification entered by the user for this region (e.g., benign, pathogenic, unknown).
ACMG Interpretation	
Audit Log	Records all actions such as changing the default transcript and the classification.
Notes	Any notes the user wants to enter. This also by default states manually altered if the user added a call.
Variant Interpretation	
Relevant Genes	
Pre-Classification Comments	Rule(s) used to pre-classify the event using an applied decision tree.
Genes	Genes (listed as symbols) in this region.
Gene Count	Number of genes in this region.
Phenotypes	Phenotypes via HPO that overlap this region.

SAP Score	Significance Associated Phenotype score - A statistical measure comparing phenotypes (HPO terms) associated with genes in a region to the sample phenotypes. Smaller scores indicate greater significance. Requires sample have associated phenotypes.
SAP Comment	Shows how the relevant genes were identified based on HPO terms associated with the gene and the patient phenotypes. The HPO terms and distance from the patient phenotype is indicated.
Similar Previous Cases	Number and percent of previous cases (including current case) with the same event. The calculation is performed for all events except for allelic imbalance events (field will be blank).
ISCN Nomenclature	Official 2016 ISCN term describing this event.
Parent of Origin	Parental source of the affected allele events if at least one parent is available and linked to Proband.
Inheritance Comments	Lists matching inheritance models for the event.

Table 6. CN and Allelic Events.

Column	Description
Min Region	The region encompassed by the midpoints of the two most external probes in the segment.
Min Length	Length in bp of the Min Region.
Max Region	The region encompassed by the midpoints of the closest probes on either end of the segment that are not part of the segment.
Max Length	Length in bp of the Max Region.
No of Probes	Number of probes in the segment.
Probe Median	Median value of the probes in the segment.
Aberrant Cell Fraction %	Using log R and BAF (when available) estimation of % aberrant cells for mosaic samples.

Estimated Copy Number	An estimated copy number for CN events calculated using the log ratio value.
Mosaic	Values Yes, No, and empty to indicate whether event is mosaic. Automatically filled if feature is turned on in processing settings. Can be manually set while in Edit mode.
BAF Segment Value	Median value of the BAF probes in the segment.
% Heterozygous	Percentage of probes lying outside the Homozygous Value Threshold – yellow lines in the plot. Applicable only to SNP arrays and CN/AOH calls from NGS data.
B/P Genes	Breakpoint genes (genes that are only partially covered by the region – possible fusion sites).
Call PValue	Significance of obtaining this call at this location (one-tailed z-test) - the probability of obtaining the observed mean of the probes encompassing the call segment assuming the true mean is zero and the distribution is normal. The value is corrected for multiple testing. If the p-value cannot be calculated for a call (e.g., for a sex chromosome), the value here will be NA.
% of DGV Overlap	Percentage of this region covered with events in DGV.
DGV Score	Score indicating similarity of an event to that in DGV. The score is a combination of similarity and number of reported cases per publication. E.g., a score of 0.88 can be achieved by perfect similarity of three cases or similarity of 88% with eighty or more cases.
DGV Score Comment	Provides the number of similar cases in DGV and the percent similarity to the cases.
VCF filter values	

Table 7. Sequence Variant Events.

Column	Description
Filter Label	Indicates which filter labels were applied during processing. Based on filter labels added to the VCF Filter Label section in the Processing settings. E.g., PASS.
Quality	QUAL column in VCF files. Phred-scaled probability for the alternate allele assertion. Higher values mean more confidence in call.
Variant Read Fraction (%)	Percent of reads supporting the Alternate Allele in this position.
Total Depth	Count of filtered reads supporting each of the reported alleles; depth of coverage.
Allele Depth	Count of unfiltered reads supporting a given allele.
Genotype	Values are homozygous or heterozygous.
Ancestral Allele	SNP allele as found in the chimpanzee.
Ref Allele	Nucleotide base on the NCBI reference assembly.
Alt Allele	Nucleotide base if different from Ref Allele.
PhyloP Score	The score measures evolutionary conservation at individual alignment sites. Useful to evaluate signatures of selection at specific nucleotides or classes of nucleotides (e.g., third codon positions, or first positions of miRNA target sites).
dbSNP	dbSNP ID hyperlinked to window providing more details from the **dbSNP** database.
GMAF %	Global minor allele frequency from one of the following data sources in order of priority: TOPMED, gnomAD Genome, gnomAD Exome, 1000Genomes. E.g., if all four sources are available, the TOPMED value will be displayed; if only gnomAD and 1000Genomes are available, the gnomAD value will be displayed. Only displayed for SNV or other variants with length no greater than one.
ClinVar	Classification and star rating from ClinVar linked to pop up window with additional details.
COSMIC	Number of records found in COSMIC linked to pop up window with additional details.

Regulatory Feature	If regulatory information is available, hyperlinked regulatory_region_variant will be displayed which opens a new window with details on the feature. If no information is available, no_consequence is displayed.
Pop. Allele Freq %	Allele frequencies from different projects hyperlinked to a window displaying additional details on each project/population. The field lists the highest frequency with project name and population. E.g., 0.865 (1KG, AMR). This means the 1000 Genomes project showed the largest frequency with Mixed American Population at 0.865 frequency. Clicking on the hyperlink opens a window displaying all projects and population frequencies.

Table 8. Transcripts.

Column	Description
Transcript Overview	Provides hyperlinked consequence value which opens a window listing all RefSeq and Ensemble transcripts as well as additional details for each. The consequence listed here is the one that is most interesting (most severe). Bold indicates canonical transcript.
Transcript ID	Accession number of the canonical transcript or manually selected transcript.
HGNC	HGNC gene identifier.
HGVSc	HGVS coding sequence name.
HGVSp	HGVS protein sequence name.
Codons	If the variant is in the coding region, the reference and alternate codons are displayed with the variant base(s) highlighted in upper case.
Consequence	Most severe consequence associated with the selected transcript.
PolyPhen-2 Score	PolyPhen score.
PolyPhen-2 Prediction	PolyPhen Prediction + transcript ID.
SIFT Score	SIFT score.
SIFT Prediction	SIFT prediction + transcript ID.

Protein ID	Refseq/Ensembl protein ID linked to the respective site.
Amino Acids	If variant affects the protein coding sequence, the change in amino acid resulting from the variant; indicated with the single letter AA symbol.
Protein Position	Amino acid position in the protein.
CDS Position	Base position based on the coding sequence.
cDNA Position	Base position of the variant based on the cDNA sequence.
Exons	Displays the affected exon (exon number) out of the total exons in this transcript (e.g., 6/10 means that the variant affects exon 6 and there are a total of ten exons in this transcript).

Table 9. In-Silico Predictions.

Column	Description
Mutation Assessor	Prediction + score
Meta SVM	Prediction + score
MetaLR	Prediction + score
FATHMM Score	FATHMM score
FATHMM Prediction	FATHMM Prediction + transcript ID

Table 10. Affymetrix OSCHP.
Only for Affymetrix OncoScan OSCHP data

Column	Description
TuScan Total Copy Number	Displays number of copies in this region where available.

Table 11. Structural Variant Events.

Column	Description
Fusion Junction 1	Displays break end region(s) for fusion junction 1
Fusion Junction 2	Displays break end region(s) for fusion junction 2
SV Quality	Displays SV Quality score
Molecule Count	The number of molecules that support the event
VCF filter values	Displays annotations by Solve for the SV Event as “PASS”, “Low Confidence”, “Masked” and “Poor Molecule Support”
% in OGM Control DB	Displays the frequency in percent of the SV event in the OGM Control DB
SV VAF	Indicates the variant allele frequency
Zygosity	Indicates the zygosity of the SV event. It is listed as “homozygous”, “heterozygous”, or “hemizygous”, but only for insertions, deletions, translocations, and inversions.

Significance Associated with Phenotype (SAP) Score

The phenotypes associated with all the genes in an aberrant region are compared with the list of sample phenotypes using a statistical measure to arrive at the SAP score.

The significance of a gene based on the sample phenotype considering all phenotypes (including super classes) associated with a gene and all phenotypes associated with a sample is determined. Phenotype weights are also considered to generate the score. A significance value is calculated using the Fisher’s Exact Test and is averaged across multiple genes overlapping an event resulting in the SAP Score.

The **SAP Score Comment** column indicates how the relevant genes were identified based on the HPO terms associated with the genes and their distances from the sample phenotypes. The column will display the phenotypes grouped by levels as described in the section Phenotype-based gene panels.

Data Export and Reporting

Export: This button will export data in visible columns for table export plus information about the sample (from the **Sample Info** window) into a tab delimited text file. The user will be prompted to select the location for the file. The default name provided is displayed; if there is no display name specified, then the sample name is used. The file name can be changed during export.

Export the Events table from within an open sample: The event table for each sample can be exported as a separate tab delimited txt file when the sample is open for review, as shown in **Figure 118**. The user selects the events and the information to export.

NOTE: When none of the events have the **Select** box checked, then all events will be exported.

Only the columns displayed in the table will be exported. To export additional data, use the **Table Preferences** icon to add additional columns to the **Table** view and then export the data. All rows will be exported unless the **Select** column is visible; in this case, only those rows that are checked will be exported.

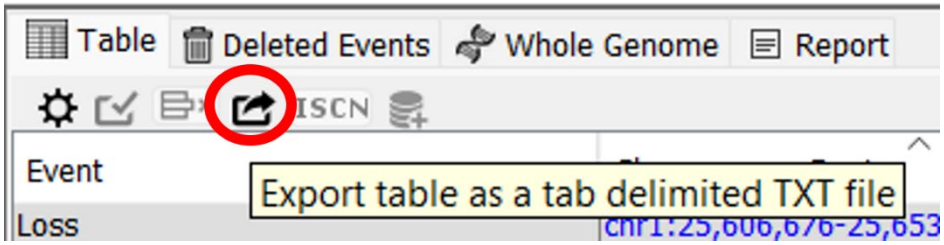


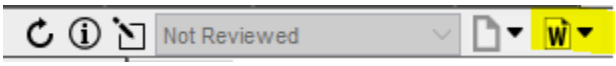
Figure 118. Export from the events table.

Exporting a Report from Query Results: Locked samples will have an icon with three bars next to the sample name. Clicking the icon will directly export the report for this case from the **Home** page. The sample does not need to be opened to generate the report.

Word Report Generation

Another option for data export is to output it into a Word document. The Admin can create and upload various Word templates through the Admin interface. The user can then select one of these templates to output results into a Word document. Tags are added to the Word template using the MS Word merge field feature.

Generating a Word Report: Click on the **Word Report** icon in the toolbar (highlighted in yellow below) and select a report template from the dropdown menu:



After selecting a template, a file chooser opens with the **File Name** field pre-populated with the **Sample Name**. The file name can be edited as well as the folder location. The report will be saved as a .docx file.

Adding a Word report template requires the user privilege Ability to perform admin operations and the user must be logged in with Admin privileges to add a report template.

In the **Sample Types Reports->Word Template** section, use the icons to add, remove, or rename a report template. A table displays the tags used in the template selected in the dropdown, as shown in **Figure 119**.

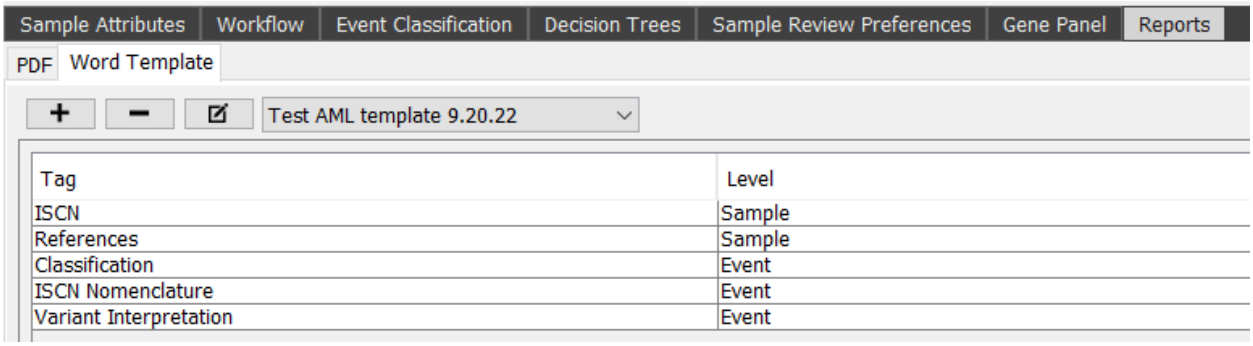


Figure 119. Template selection.

Exporting Sample Query Results: Sample details displayed for samples queried on **Home** page can be exported into a text file using the **Export** tool under the **Samples** menu, seen in **Figure 120**.

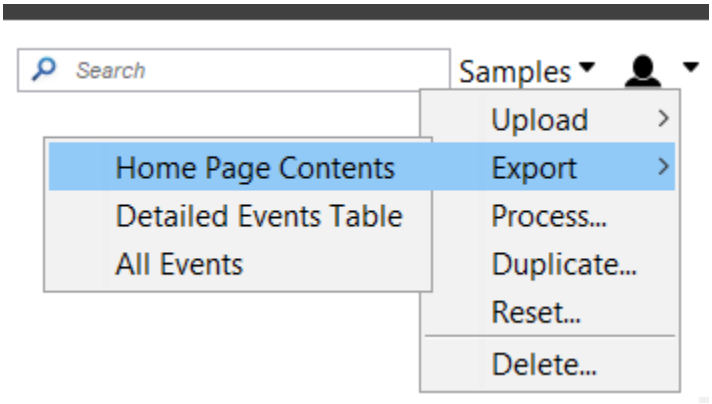


Figure 120. Exporting query results.

Export events table for a batch of samples: The event table for each sample can also be exported as a separate tab delimited txt file for each sample through the **Home** page. To export sample data through this route, each of the samples to export will need to be in the locked state. The user can then query the list of samples for export and select **Samples > Export > Detailed Events Table**.

NOTE: The exported table will export only the selected events and detailed information at the time the sample was locked.

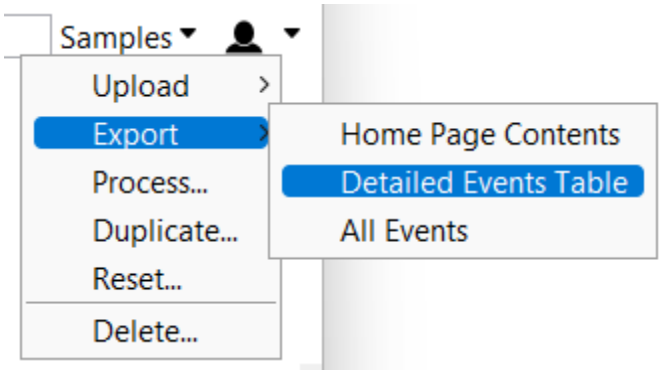
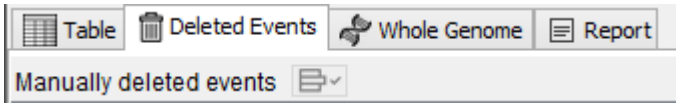


Figure 121. Export events table.

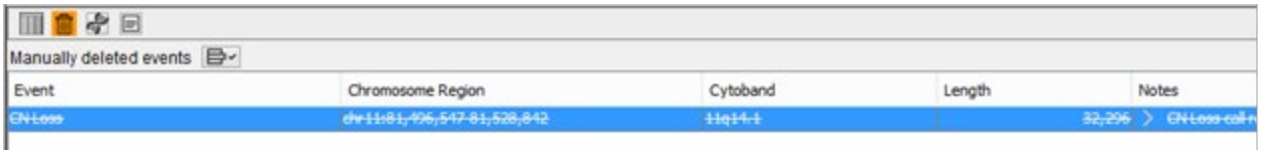
Manually Deleted Events

Any events that have been manually deleted are shown by clicking on the **Deleted Events** tab.



If events have been deleted the icon will turn orange and the events will be listed in the table. Selecting an event, or all events, and Clicking on the **Restore** button will restore the deleted event/s. If the **Audit Log** column is displayed in the table, it will show who deleted the event and when for the deleted events.

NOTE: for events to be deleted or restored the sample must be in edit mode. Not all users may have been given the privilege ability to delete samples.



Variant Details Tab

The **Variant Details** tab collates all essential information about a variant in a single easy-to-read layout that automatically adjusts to the type of variant being reviewed (e.g., **CNV** vs **Seq Var**) as well as Test Type (Oncology vs Constitutional). The sample Test Type, **Constitutional** or **Oncology**, will dictate the layout of the **Variant Details** tab and the type of information being displayed, as shown in **Figure 122**.

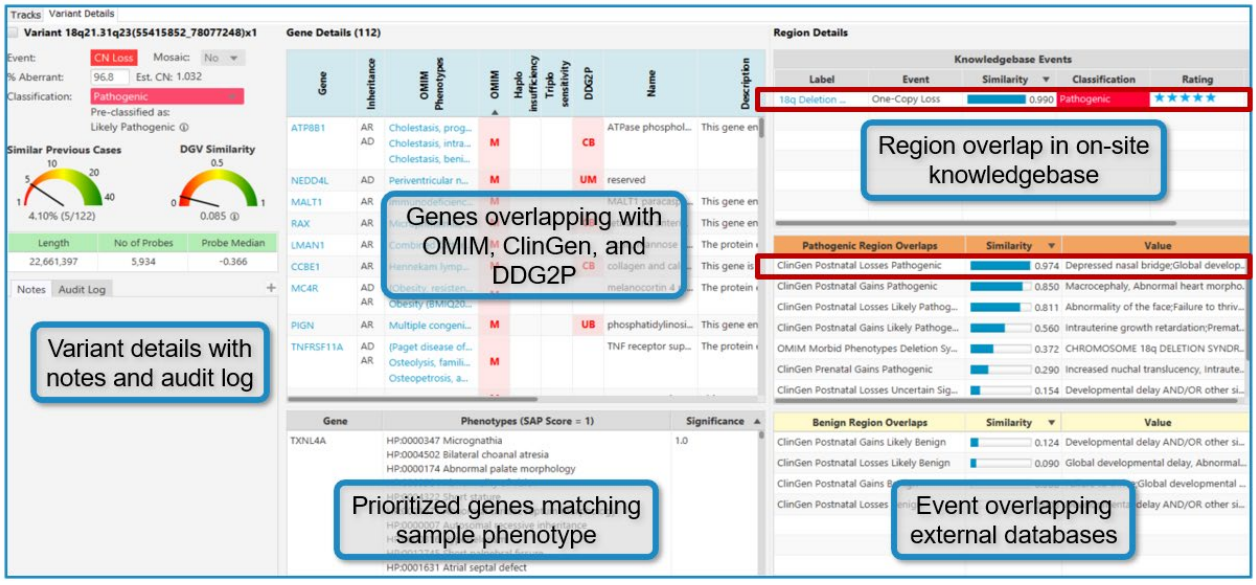


Figure 122. Variant Details layout.

The Whole Genome View

Selecting the **Whole Genome** tab displays the data as a whole genome view for copy number (or Log2Ratio) plot and BAF plot. This view is ideal for confirming or determining the gender of the sample and for giving an overview of the data, as seen in **Figure 123**.

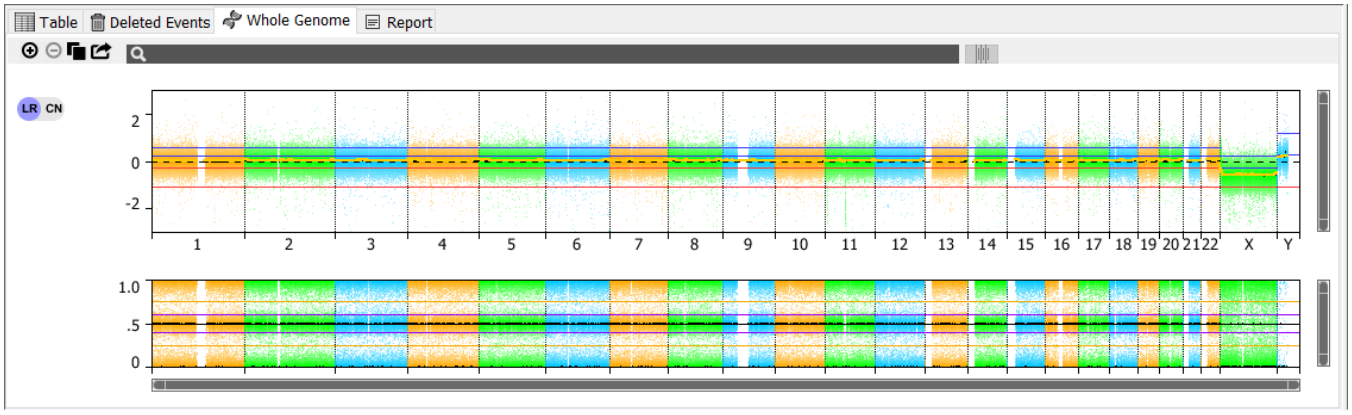


Figure 123. The whole genome view.

Tools in **Whole Genome** tab include:

- **Zoom in/out** horizontally.
- **Copy:** Clicking this tool will copy the probes plot to the clipboard; the contents can then be pasted into another application.
- **Export:** Clicking this will export the whole genome plot to be saved as a png or jpg file. A save dialog will ask for the folder to save the picture, an option to select png or jpg, and the option to open the file after saving is complete, shown in **Figure 124**.

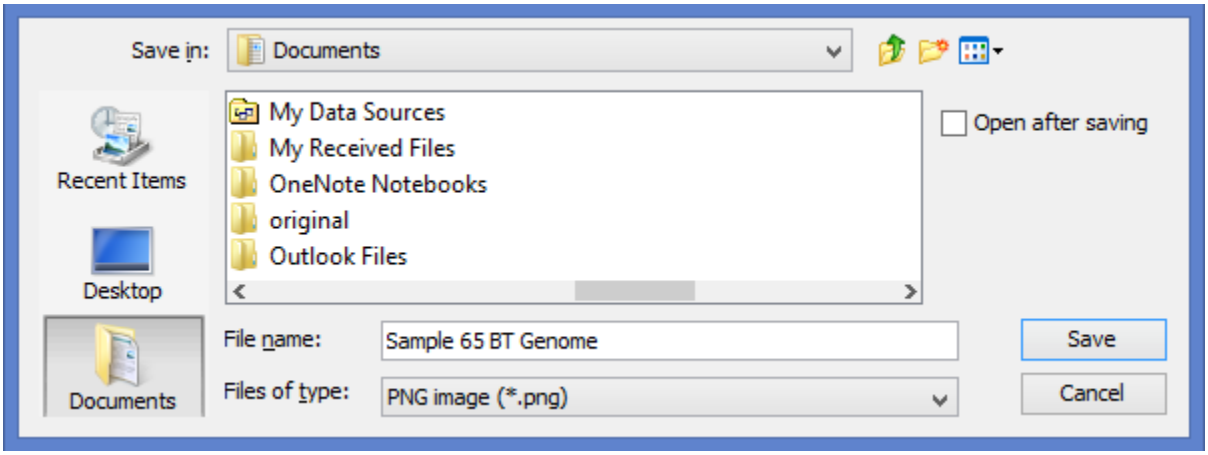


Figure 124. Export tool.

The zoom icons allow for zooming along the X-axis on the plots (probes, BAF). Plots can be zoomed along the Y-axis by using the mouse and keyboard keys: press and hold down the Ctrl key while moving the mouse scroll wheel to zoom in/out or via the slider bars to the right of the plots, as seen in **Figure 125**.

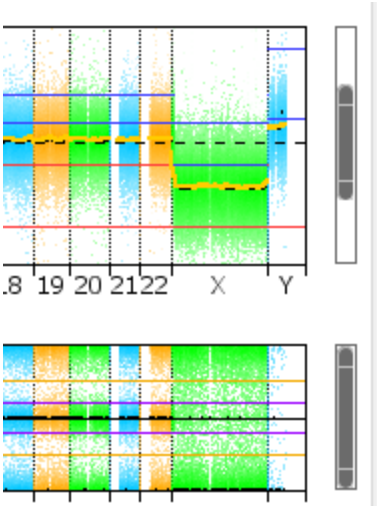


Figure 125. Slider bars for the zooming tool.

The Report View

Clicking on the **Report** tab for the first time after a sample is opened will generate a report of the variants visible in the table. Depending on the number of events in the table, creation of the report could take several minutes – a note in the window will indicate this until the report is finished and when finished, the variant details will be displayed. Items displayed in the report are dependent on the columns visible in the **Table** view.

The report describes all aberrations found in the sample after filters have been applied, as shown in **Figure 126**. Information from this report can easily be copied and pasted into external reports. Contents of the report are displayed in chunks (approximately forty events per page). The bottom of the section has page numbers that are hyperlinked to easily jump from one page to another.

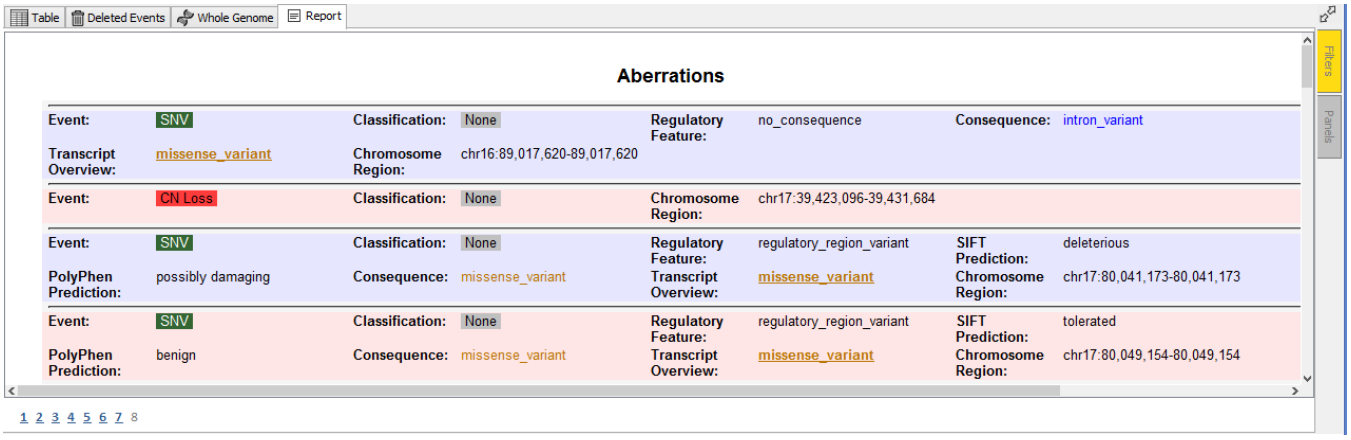


Figure 126. Aberrations found in the Report view.

Event and **Classification** values are highlighted in the same color matching the relevant event type or classification. Clicking on the event zooms in on the event in the genome browser. Annotations such as consequences and **ClinVar** classifications are also displayed in the same color coding as that in the table. Some fields (e.g., **transcript overview**, **OMIM** and **Morbid Phenotypes**) are hyperlinked to a window with details, shown in **Figure 127**.

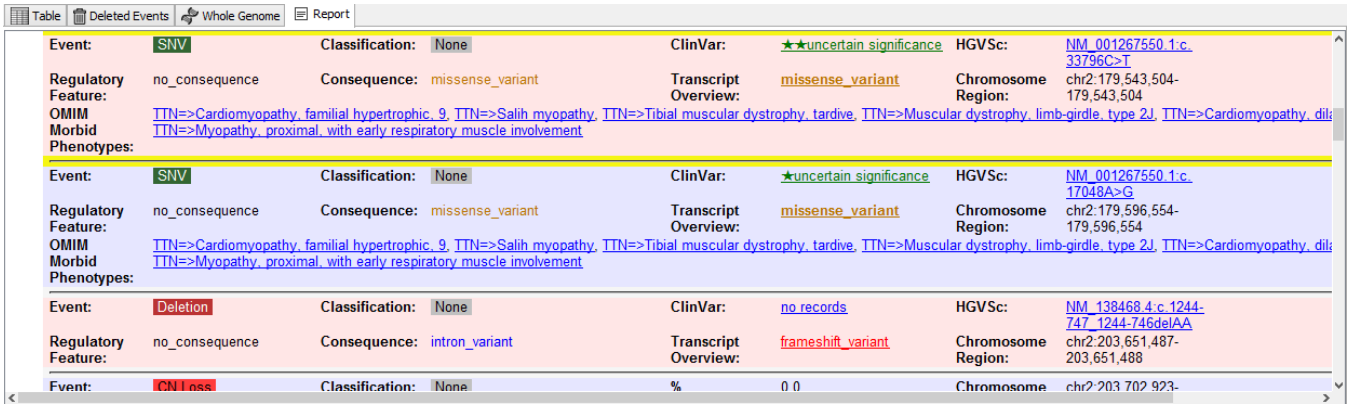


Figure 127. The event row selected in the table is highlighted in the Report by yellow lines.

Gene Panel Selection/Import

If the sample type has gene panels associated with it (created by the Admin), then the **Panels** tab will list the genes/regions in the selected panel. At minimum genes and/or regions will be specified. Specific transcripts for genes as well as a minimum read depth can also be specified, covered in the section below. Basic panel features are shown in **Figure 128**.

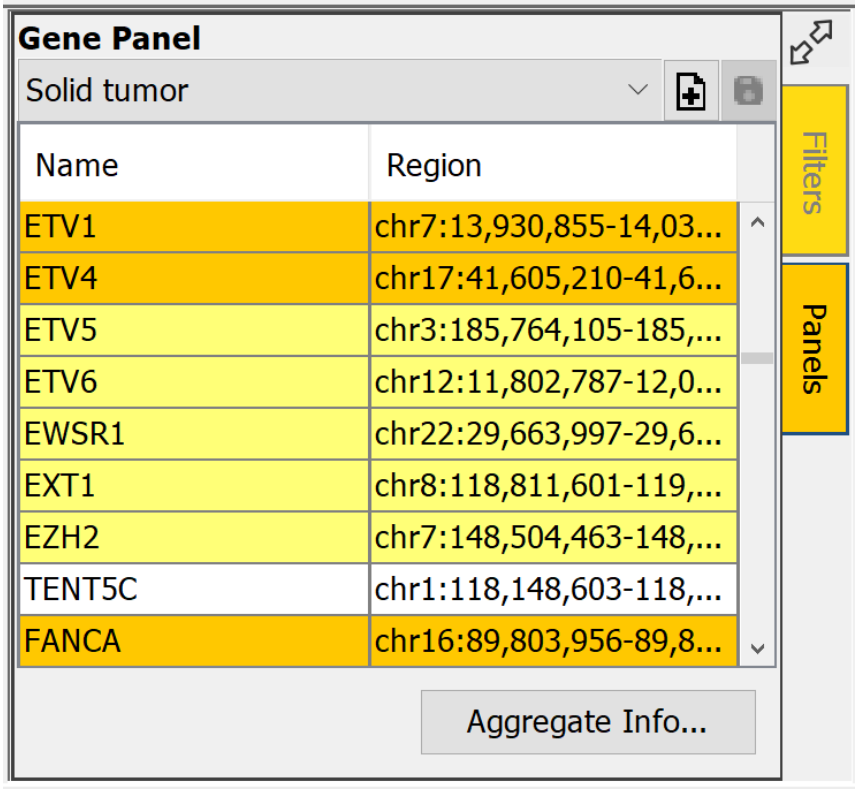


Figure 128. Basic gene panel features.

The default sort order of genes/regions will be the order in the upload list or the order in which the genes/regions were entered manually by the Admin.

This list can be sorted alphabetically (genes) or by chromosome number/base position (regions) by clicking on the header. The sort order cycles through unsorted, ascending, and descending with multiple clicks of the header. A blue triangle pointing up on the header indicates ascending or descending (triangle pointing down), shown in **Figure 129**. If there is no triangle, the list is not sorted but rather in the original order as entered by the Admin.

Ascending **Descending**

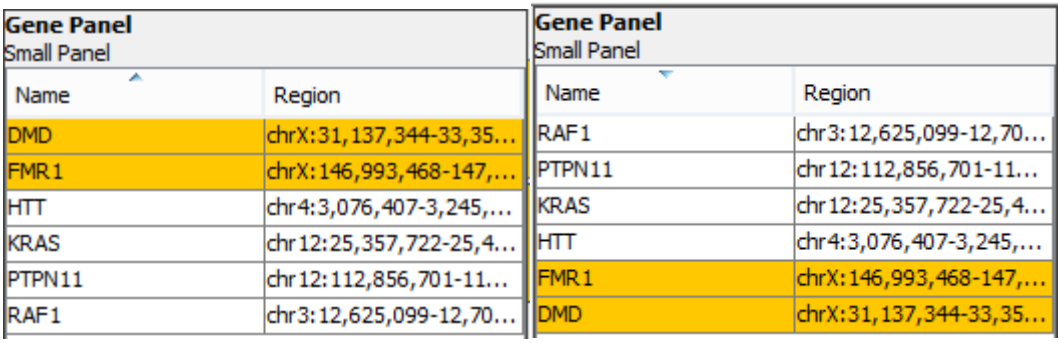


Figure 129. Ascending and descending ordering of gene panel entries.

A specific panel can be selected via the **Filters** tab in the **Panel Selection** filter by clicking on the **Gear** icon, which opens the **Filter Parameters** window. Here, a panel can be selected from the dropdown list, as seen in **Figure 130**.

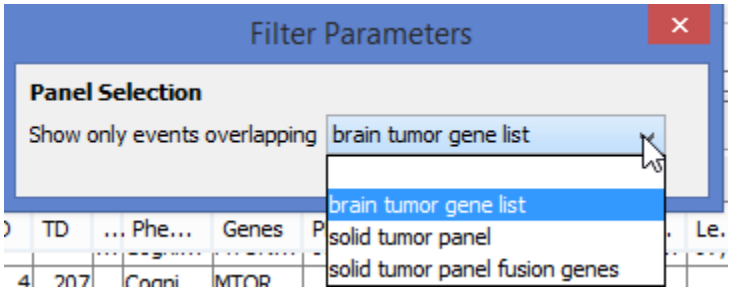


Figure 130. Filter Parameters selection.

After a panel is selected, its name will appear in the filter box and a checkmark will be displayed in the box. In some cases, users may not be able to select the panel. This is due to a lock on a selected panel with a sample type by the System Admin. In such cases, the user cannot check/uncheck the box. The checkbox will be grayed out and already checked, as in **Figure 131**, with the Admin selected panel displayed, as shown in **Figure 132**.

Locking Panel Selection: The Administrator can lock a selected panel with a sample type so that only the selected panel will be applied to all samples of that type during sample review. Users will not be able to select any other panel, add phenotype-based panels, or add ad-hoc panels. The panel can be locked in the **Sample Review Preferences - The Filter tab** section. Locking of the **Panel** prevents the user from seeing events other than those on the **Panel**. The Admin may do this to prevent incidental findings as specific samples may only be allowed evaluation only for certain genes.

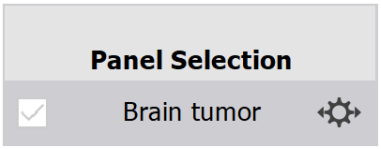


Figure 131. A locked selection panel indicator.

Shading of the **Name** and **Region** columns in the panel list indicates the status of the gene:

White – gene has no genomic variants identified.
Yellow – gene does overlap with a variant, but the filters applied are removing the variant(s).
Gold – gene has visible variant(s)

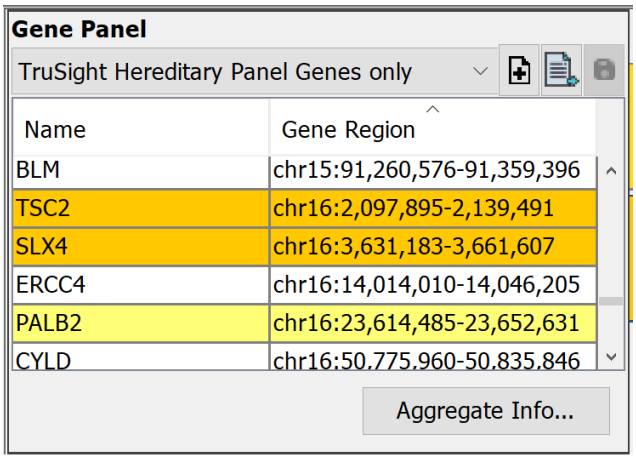


Figure 132. Admin selected panel.

If no filters are applied, the same panel shown above will be seen as in **Figure 133** (yellow highlighting is gone).

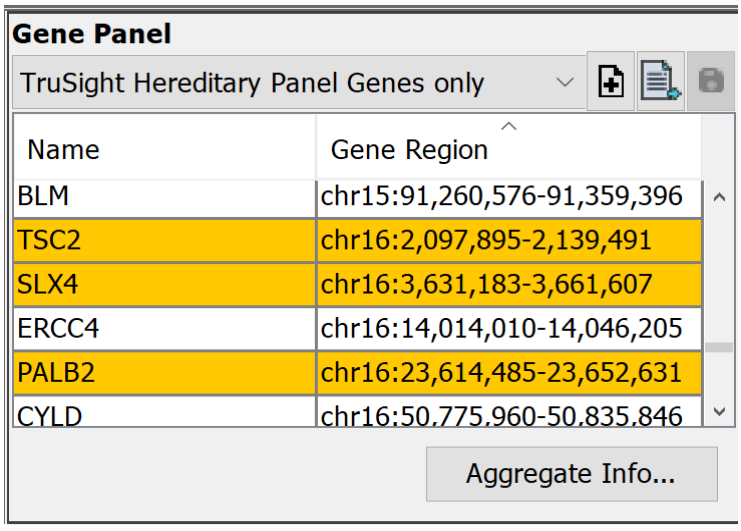


Figure 133. No filters applied.

Clicking the **Aggregate Info** button at the bottom of the panel list expands the section to reveal aggregate information on the selected panel as number of regions in each category listed below.

- Total panel regions
- Panel regions with events [yellow highlighted]
- Panel regions with post-filter events [gold highlighted]

Keyboard triangle keys can be used to step through each gene on the panel list for manual inspection. The genome browser will zoom in on the gene selected in the panel and the event row(s) will be highlighted in blue in the table. In **Figure 134**, BRCA1 is selected in the panel and the browser is zoomed in on the region of the gene only.

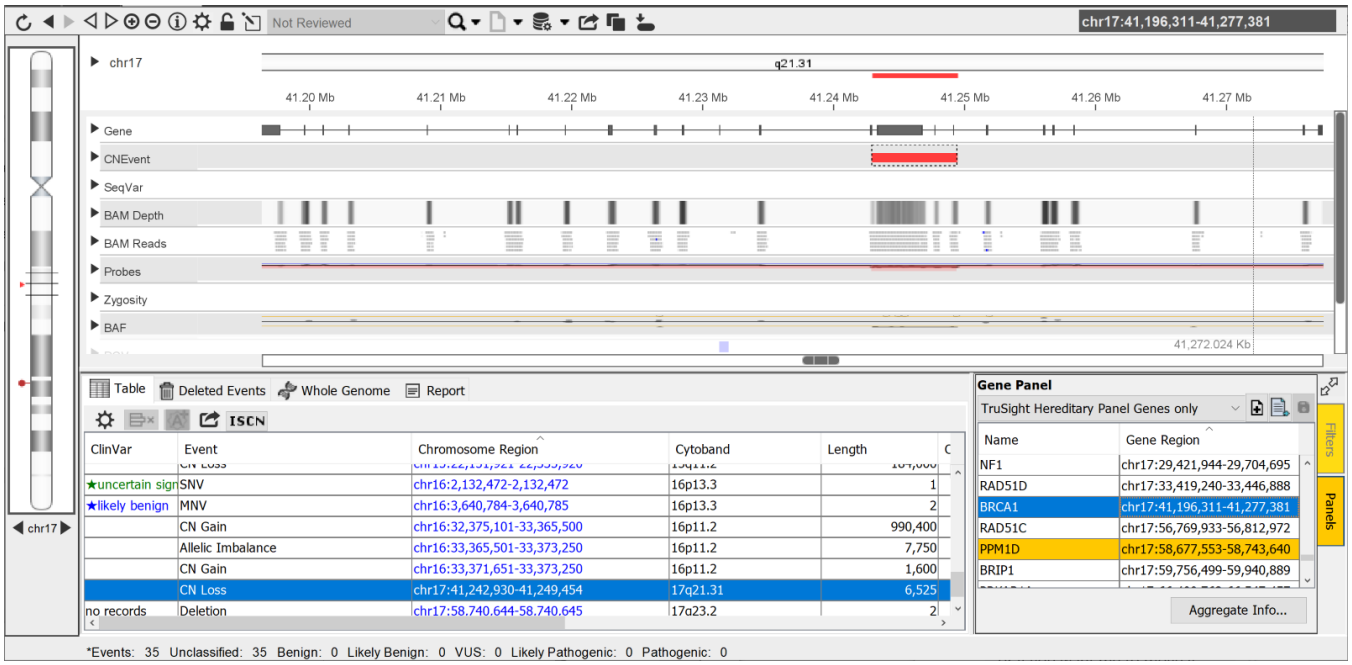


Figure 134. BRCA1 selection and zoom.

VALIDATION OF A GENE PANEL

When a panel is first loaded, the genes/regions are validated against the current annotations and the regions of the genes are saved. This allows the software to later know where the gene mapped when it was first loaded as the positions could change with annotation updates. In versions prior to 5.0, the regions were not validated and saved. If a panel that was loaded in a prior version is applied in later versions of the software, a warning will be displayed if the software finds that some locations of genes have changed. The details are displayed as in **Figure 135**.

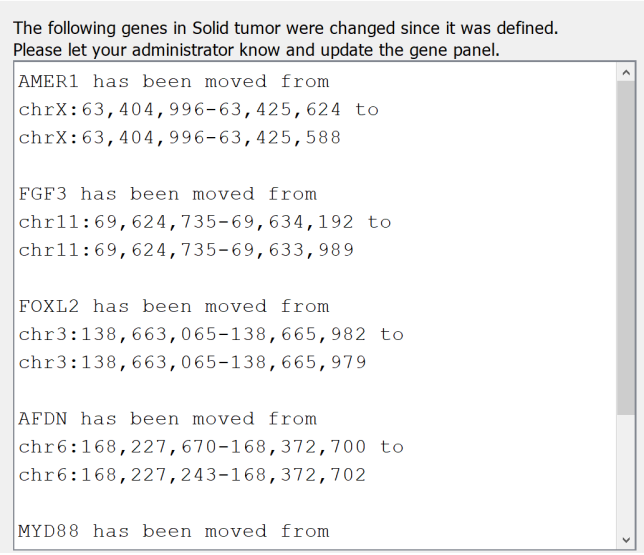


Figure 135. Warning that locations of genes have changed.

If the **Panel** was loaded by the user, the user can update the regions. If the panel was uploaded by the Admin, then the Admin will need to update the panels to validate them and save the regions.

IMPORTING A GENE PANEL FOR A SPECIFIC CASE

See section on *Importing Panels for Sample Types* within the Administrator section.

The user can also import a gene panel to be applied only to a specific sample (the one under review). Use the **Load a temporary gene panel** button to load a panel as in **Figure 136**.

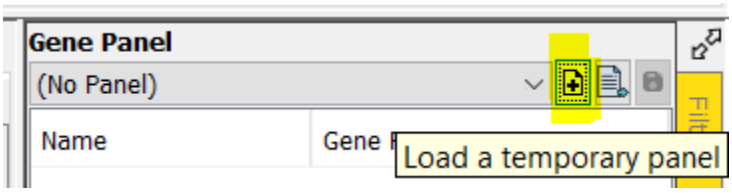


Figure 136. Loading a temporary gene panel.

A file chooser will pop up allowing selection of a CSV, TSV. or BED file. Please review the Admin creation of gene panels section.

SAVING A GENE PANEL

One gene panel can be saved with the sample using the **Save** button indicated in **Figure 137** below. This is particularly useful for ad-hoc/temporary panels loaded by the user which are not already associated with a sample type. The next time the sample is opened, the ad-hoc panel will be available for that sample. Ad-hoc panel that is not saved using this tool will not be available the next time the sample is opened.

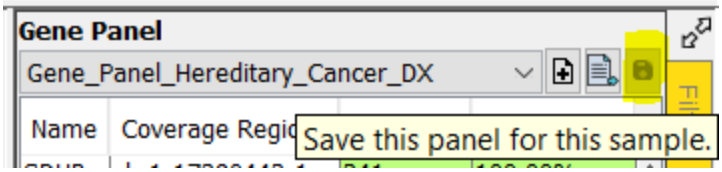


Figure 137. Saving a gene panel for a sample.

EXPORTING A GENE PANEL

The data in the gene panel can be exported and saved as a tab-delimited text file using the **Export** button indicated in **Figure 138** below.

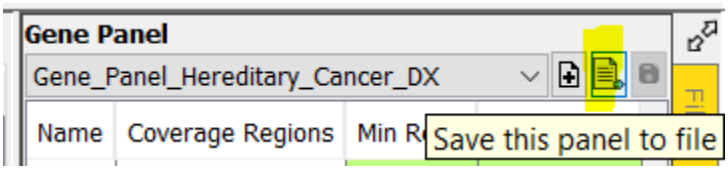


Figure 138. Exporting a gene panel.

TRANSCRIPTS AND COVERAGE THRESHOLDS FOR PANELS

To avoid potential false positive results due to poorly covered regions, coverage thresholds (and transcripts) can be specified during panel creation. The Admin may specify one or more transcripts (displayed in the **Coverage Regions** column) and a coverage threshold (displayed in the pop up when hovering over the **Min Read** column) for each gene. In such cases, the other columns (**Min Read**, **Coverage %)** in the **Sample Review** window will have values and will be color coded as indicated in **Figure 139**.

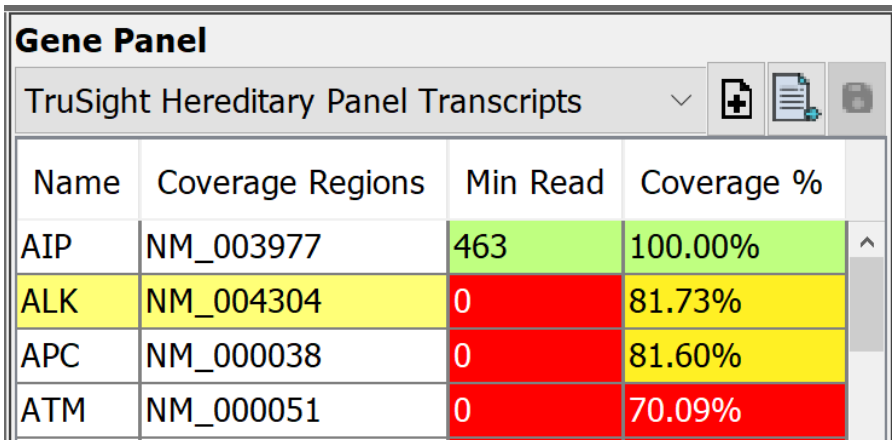


Figure 139. Color-coded values.

If transcripts are specified, the coverage region is taken as the union of all exons in the specified transcripts. If a transcript is not specified, the region is taken as the entire genic region including introns.

NOTE: The minimum read depth for the gene panel is counted using all reads whereas the read depth track in the browser displays filtered reads (excluding PCR duplicates and secondary alignments).

- **Min Read:** The observed minimum reads in this region. Hovering over the field displays a pale-yellow pop-up box with the observed minimum reads and the specified minimum read depth.
- **Coverage %:** Displays how much of the region meets the specified coverage threshold (Min Read Depth).

Hovering over the **Min Read** column values will display the min read count and coverage threshold as a ratio. In **Figure 140**, the panel coverage threshold (min read depth) was specified as 100 for all genes so the pop up in the figure below displays 463/100 (min read count/coverage threshold) for gene AIP. **Figure 141** is a legend for the color-coding.

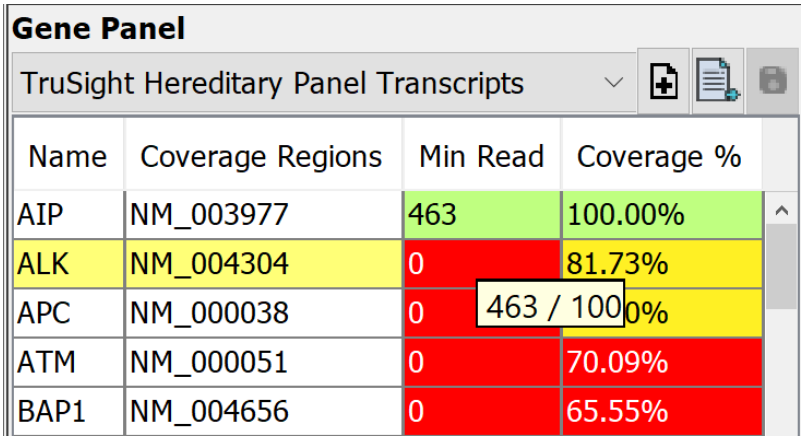


Figure 140. Min Read and Coverage % for a specific gene.

Min Read field color if minimum read count at any position in the region meets the following coverage thresholds	% Coverage color if following percentages of reads are below the coverage threshold
<90%	80%
Between 90% and 100%	Between 80% and 90%
>=100%	>90%

Figure 141. Legend for color-coding.

If one of the rows in the **Gene** Panel list is selected and therefore highlighted, as in **Figure 142**, the cell colors will be of a deeper shade than those displayed in **Figure 143** to indicate the cell is highlighted.

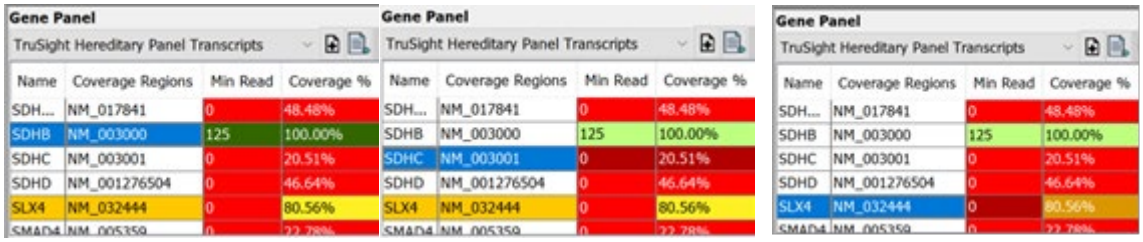


Figure 142. Selected/highlighted rows.

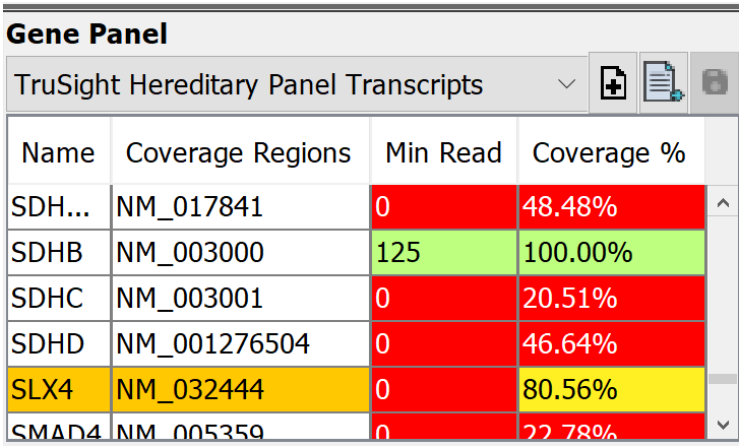


Figure 143. No row highlighted.

PHENOTYPE-BASED GENE PANELS

Whereas the Admin adds gene panels and associates them with specific sample types, the phenotype-based gene panel can be applied by the reviewer (no Admin rights needed) for an individual sample.

Addition of this panel can be accomplished via the **Filters Tab > Panel Selection** (by clicking on the **Gear** icon) which will open the **Panel Selection** window, as seen in **Figure 144**, or it can be accomplished via the **Panels** tab by clicking on the dropdown under **Gene** panel, as seen in **Figure 145**.

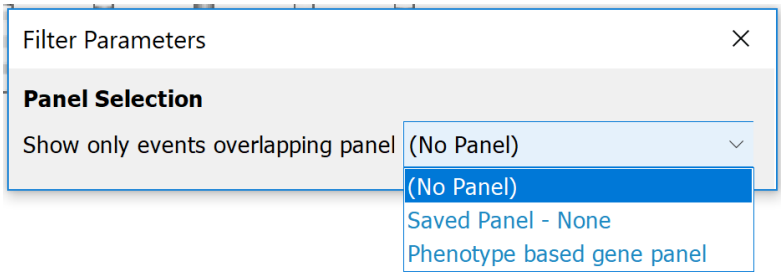


Figure 144. Panel Selection in the Filters tab.

Figure 145. Another way to add a panel.

At least one HPO term must be associated with the sample for this panel to have an effect.

As seen in **Figure 146**, before application of the phenotype-based gene panel, there are 876 events.

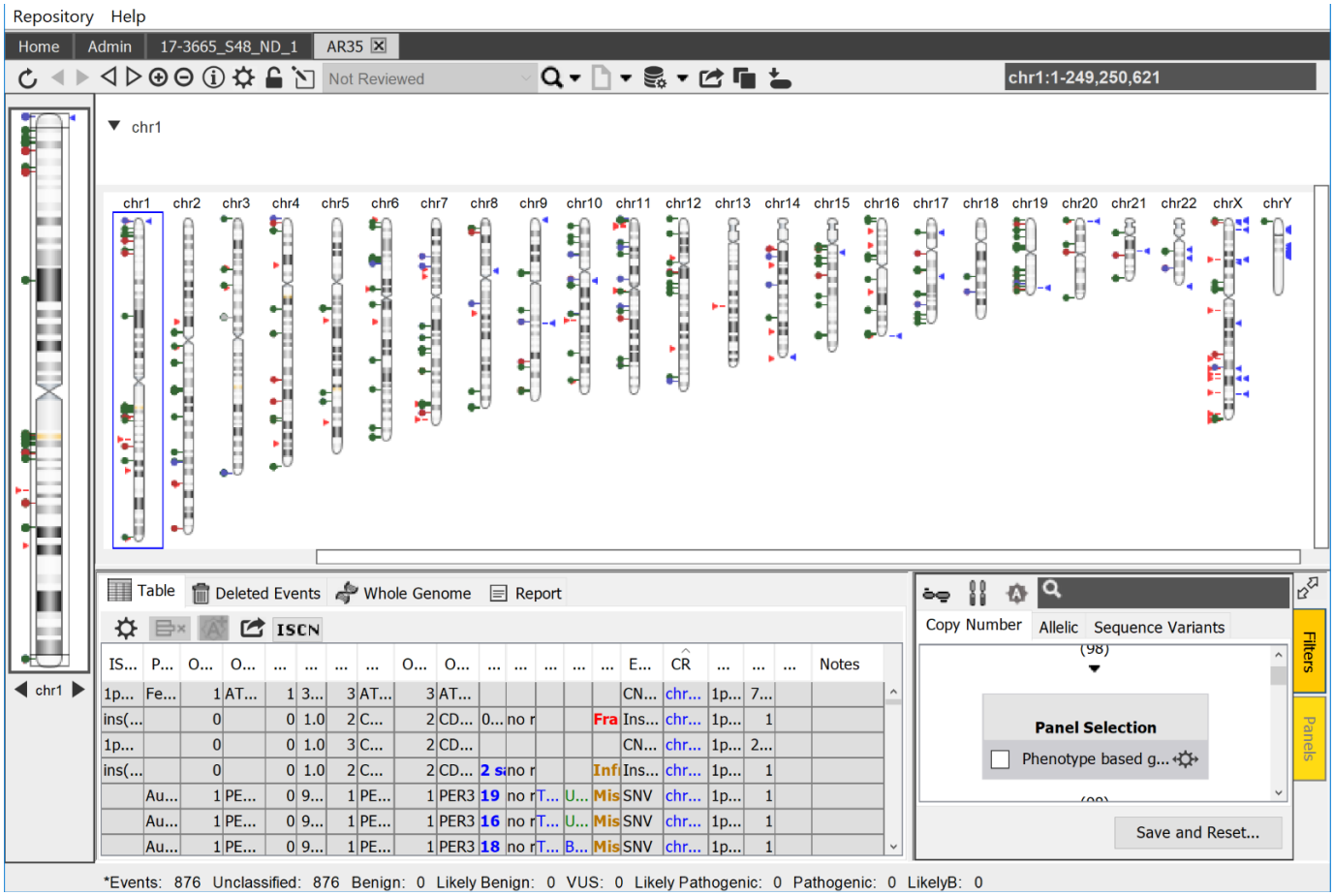


Figure 146. Before the phenotype-based gene panel was applied there were 876 events.

After application of the phenotype-based gene panel, the number of variants decreases to thirty, as seen in **Figure 147**. In the **Panels** tab, each region associated with the phenotype is listed along with a p-value, sorted with the most significant value at the top. The user can click on the column headers of any column to sort as ascending, descending, or unsorted (menu arrows on the column headers indicate sort order). To display only the most significant regions, a significance cut-off can be specified in the Max sig. threshold field (default is 0.01), shown in **Figure 148**. Specifying 1E-15 limits the list to only those regions that have a p-value less than the specified threshold of 1E-15.

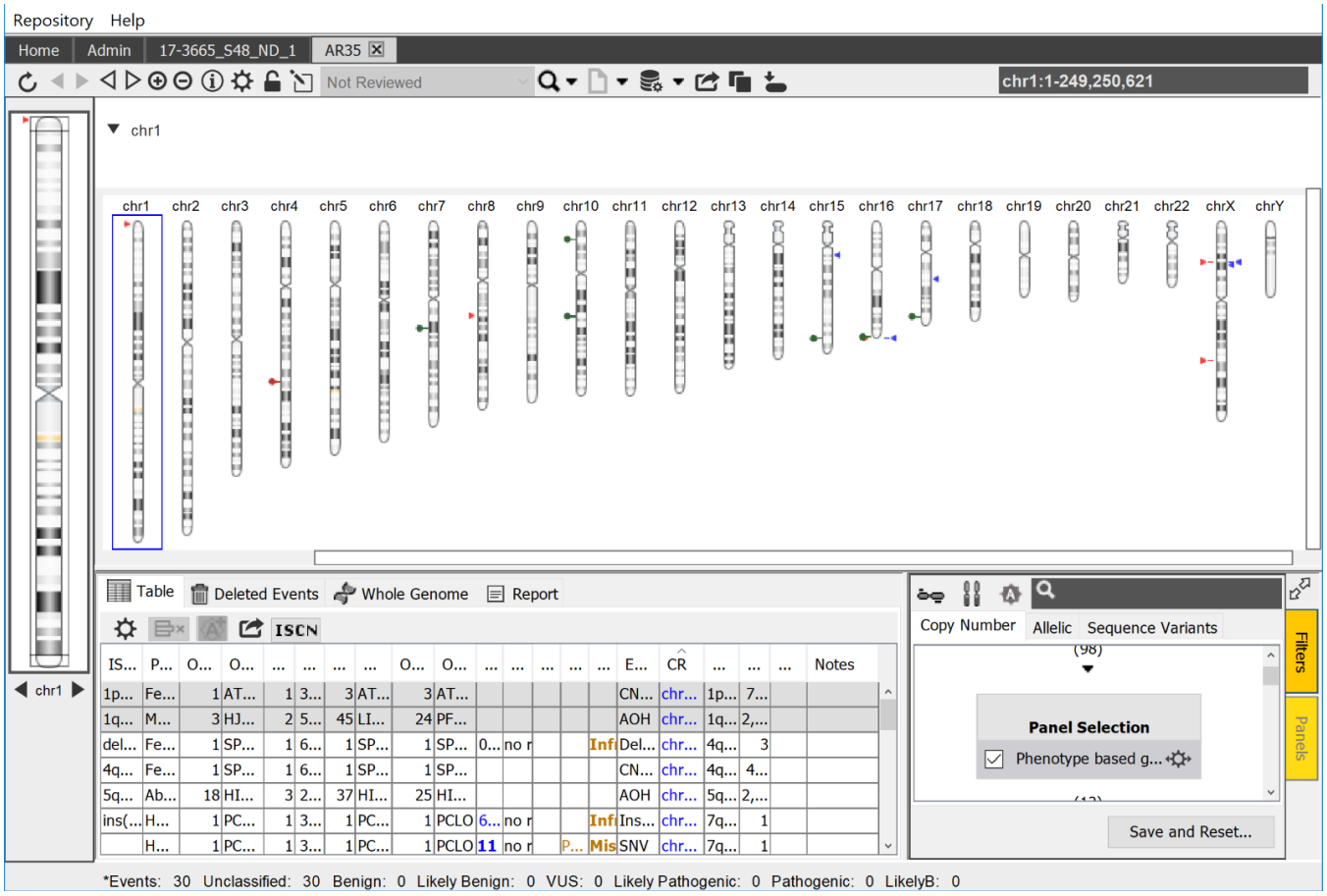


Figure 147. The number of events has decreased to thirty after applying the phenotype-based gene panel.

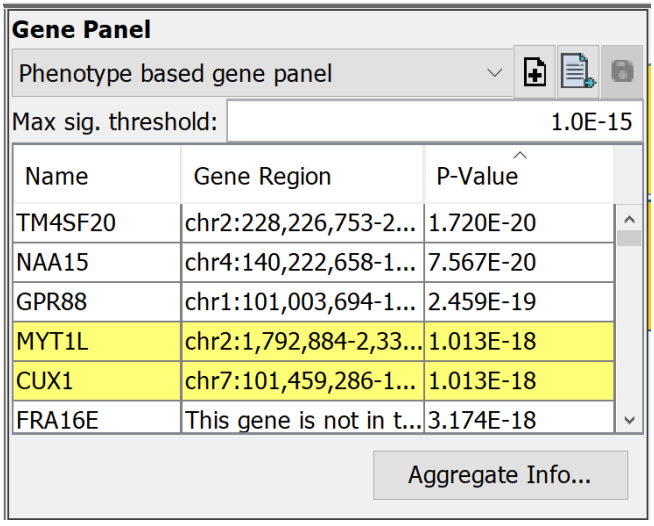


Figure 148. Specifying a maximum significance threshold for the phenotype-based gene panel.

If the HPO terms associated with the sample change, the phenotype-based gene panel is updated to reflect this change. A gene imported into a phenotype -based panel that does not exist in the current RefSeq database will display a message stating as such as there can be differences in the genes obtained via HPO versus RefSeq.

Hovering over the gene symbol in the list displays a tool tip indicating how the gene was identified by displaying which sample phenotypes link to that gene and how (to what degree). The terms are displayed in levels (based on distance from the sample phenotype) in the following format, shown in **Figure 149**.

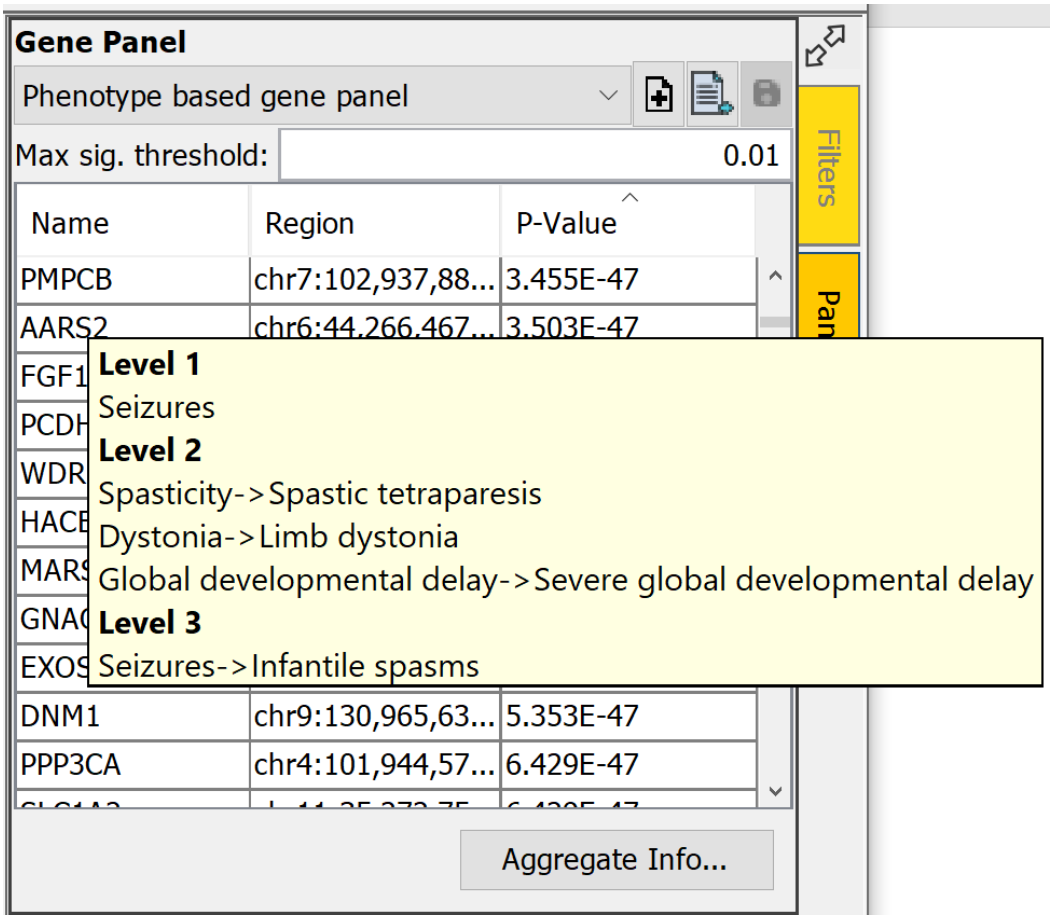


Figure 149. Gene HPO term -> patient phenotype (HPO term) linking to the gene HPO term.

The HPO terms linked to the gene are segmented by levels (up to three levels will be displayed), as described below:

- **Level 1:** Exact match to sample phenotype
- **Level 2:** One node away from sample phenotype.
- **Level 3:** Two nodes away from sample phenotype.

Figure 149 above is from a sample with phenotypes that include Seizures, Spastic tetraparesis, Limb dystonia, Severe global developmental delay, and Infantile spasms. The tooltip is for the gene AARS2.

Level 1 lists only Seizures because it is the direct sample phenotype associated with the gene AARS2.

The sample phenotype Spastic tetraparesis linked to the gene AARS2 is **Level 2** because it is one node away from the HPO term Spasticity, which is directly associated with AARS2, see in **Figure 150**.

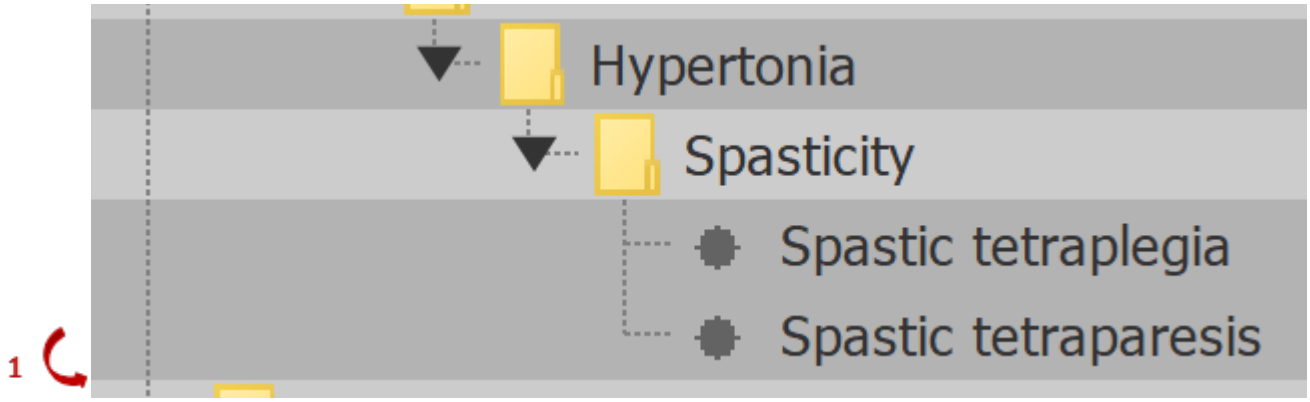


Figure 150. Sample phenotype is one node away from Spasticity.

Sample phenotype Infantile spasms linked to AARS2 is Level 3 because it is two nodes away from the AARS2 HPO term Seizures, as seen in **Figure 151**.

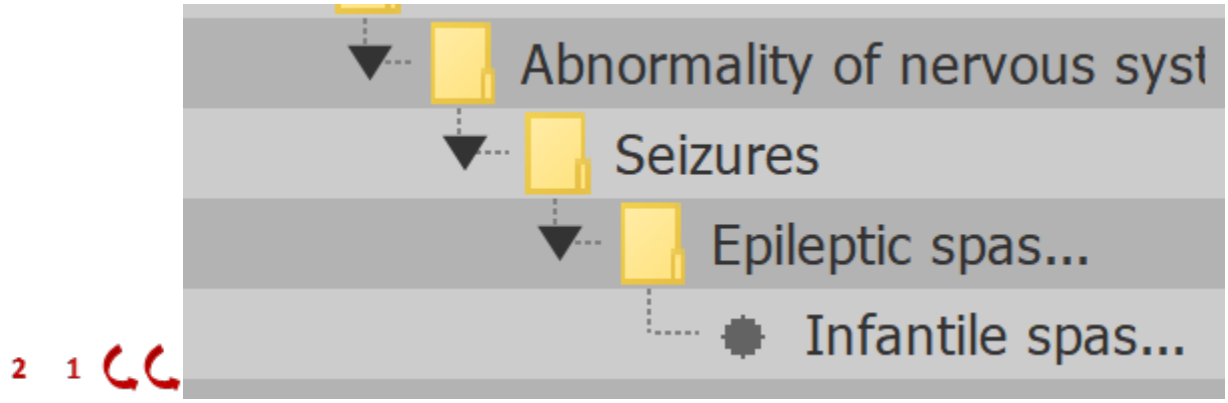


Figure 151. Two nodes away.

Guidelines for Reporting

The **Guidelines** feature is a list of variants in a table format that the user can mark as “Detected” or “Not Detected”. The primary purpose of the **Guidelines** feature is to easily transfer as a table onto a report template.

CREATING AND LOADING GUIDELINES FILE

To utilize the **Guidelines** feature, a **Guidelines** file will need to be created. This file should list the name of a variant in each row in free text format. The header (or first row) of the file should be “Guideline-based variants”. The file is saved with .txt extension.

Admin user may upload the **Guidelines** file by navigating to **Admin tab > sample types** (select the sample type name) **> Guidelines** tab and then clicking on the **+** button. The name of the .txt file will load as the name of the guideline and each row of the .txt file will be displayed as a separate variant. Only one guideline may be associated with a sample type. Once a guideline has been used by a sample in the sample type, then the guideline cannot be replaced or deleted.

GUIDELINES IN SAMPLE REVIEW

When reviewing a sample, the **Guideline Variants** list can be viewed in the **Guidelines** tab under the **Panels** tab to the right of the table panel (see **Figure 291**).

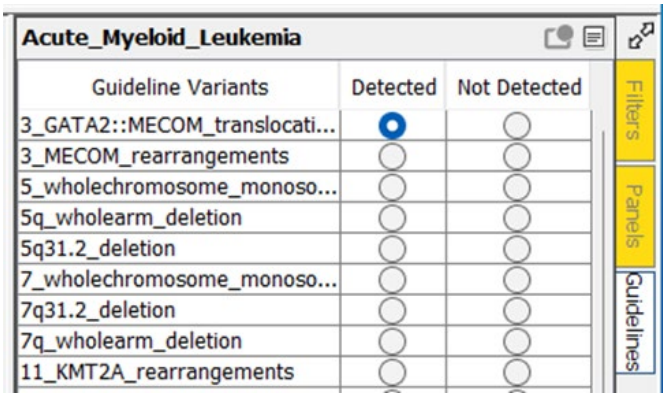


Figure 291. **Guidelines** tab.

In sample edit mode, users may mark each variant in the list as “Detected” or “Not Detected” by clicking on the circle button for each row. Additionally, users may select multiple variants by holding down the **CTRL** button and clicking on the rows. Then the user can mark as detected or not detected by clicking on the in the upper right corner of the **Guidelines** tab. Marking variants as **Detected** or **Not Detected** will be saved in the guideline-based **Variants Audit** Log. This audit log can be viewed by clicking on the in the upper right corner of the **Guidelines** tab.

Creating and Visualizing Related Samples/Trio Analysis

Related samples: Sometimes a sample is related to others in the database, and it is useful to view and analyze these in comparison to each other. One example would be samples that are from the same individual but taken at different times (e.g., for cancer samples – diagnosis, remission, relapse; or family relationships – Proband, Mother, Father, Sibling). Such samples in the database can be linked together using the **Linked Sample ID** and **Linked Sample Relationship** attributes.

To link samples, a unique ID must be assigned to the samples via the **Linked Sample ID** attribute. This ID must be the same across the samples being linked to each other . Other needed information is the **Linked Sample Relationship** field. To link samples, the samples must all map to the same genome build. Attempting to link samples belonging to different builds will display an error message.

Click on the **Sample Info** button to bring up the respective window. Fill out the appropriate **Linked Sample Relationship** details from the dropdown menu (values available here are defined by the VIA Administrator). Assign a Linked Sample ID to the sample, ensuring the same Linked Sample ID is used for all related samples. Assign a status (Affected, Unaffected, Unknown). In **Figure 152**, this sample is the Proband.

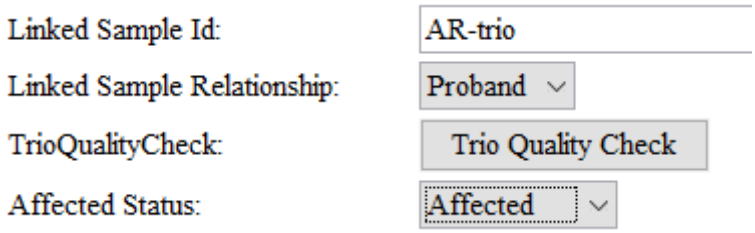


Figure 152. Proband sample.

One mother, one father, and unlimited siblings can be associated with a sample. If the **Linked Sample Relationship** is set to **Proband** the **Affected Status** is automatically set to Affected.

Trio Quality Check

ACMG guidelines recommend checking for biological family relationships when performing trio and family-based analyses. There is an option to perform trio quality checks to confirm the family relationships created for both array and NGS samples. When opening a Proband sample, VIA will warn if there is an unusually high rate of Mendelian errors (>0.01 for a single parent, >0.05 for both parents); only one parent sample is required for the function to run. Samples that can be used with VIA's inheritance pattern must be linked to other related samples, (e.g., Proband) as part of a trio. This trio quality check can also be run manually from the **Information** window .

The calculation uses the relationship meanings rather than the Linked Sample Relationship values to ascertain the relationship between samples. The trio quality check algorithm checks for Mendelian errors at positions where calls are available for the proband and all parent samples. For array samples, it checks for alleles (e.g., mother containing allele A of the proband and father allele B, or vice versa); for NGS samples, it does a similar check, but looking at the nucleotide at relevant positions. Since positions that are called as being homozygous for the reference allele are usually not specified explicitly, in practice, Mendelian errors are usually only counted at positions where all samples are either heterozygous or homozygous for an alternate allele.

After calculation is performed, the Mendelian error rate will be displayed, as seen in **Figure 153**, and state whether it is high or Ok. If the rate is high, the warning will be displayed whenever the linked sample is opened for review. For samples that have both sequence variants and array data, the error rate from both BAFs and sequence variants is reported.

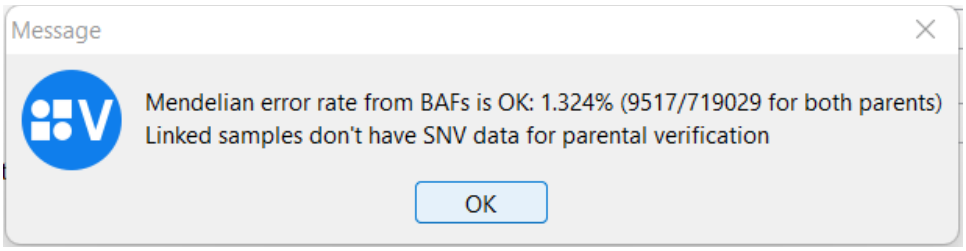


Figure 153. Mendelian error rate notice.

Parent Of Origin (Source Of Affected Allele)

Often it can be informative to know which parent was the source of the aberration. The **Parent of Origin** column displays this information for **CNV/AOH** and **SeqVar** events. For *de novo* **CN** and **AOH** events (no overlapping

events in either parent), parent of origin is determined using informative SNP probes/BAF values from the parent(s') sample(s) and the proband. The parents must both be homozygous with different alleles (AA vs. BB or vice versa) or one parent must be homozygous and the other must mostly have the other allele.

Calculations can be run if there is at least one parent linked to the proband and the family samples require SNP probes/BAF values (same processing type). If calculations cannot be run due to lack of data (e.g., proband does not have SNP probes/BAF values), the **Parent of Origin** column will be empty. If there is enough data to run calculations but BAFverification failed (e.g., the SNP probes/BAF values do not match), an error will be displayed when attempting to add the **Parent of Origin** column to the table, as seen in **Figure 154**.

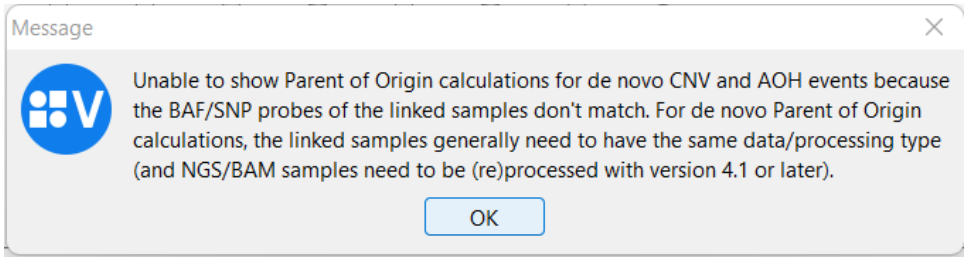


Figure 154. Parent of Origin warning message.

Identification of the parent of origin for inherited copy number events uses the similarity score between the proband and parent events. If similar events exist in both parents, it will be reported as **CONFLICTING**.

The **Parent of Origin** column from the table indicates parent of origin as well as whether the event was inherited or is *de novo*, as seen in **Table 11**.

Table 11. Parent-of-origin information.

Parent of Origin	Inheritance Comments
BOTH - Inherited	Recessive, Dominant
BOTH - Inherited	Recessive, Dominant
	De Novo, Recessive (Co...
BOTH - Inherited	Recessive, Dominant
FATHER - Inherited	Recessive (Compound ...
EITHER - Inherited	Recessive (Compound ...
FATHER - Inherited	Recessive (Compound ...
FATHER - Inherited	Recessive (Compound ...
BOTH - Inherited	Recessive, Dominant

If there are not enough probe positions to use for the calculations, a note of insufficient data will be made in the column. If the origin is identified, the parent will be noted (e.g., FATHER) and information on probe positions will be noted as well. Here are the possible values in the **Parent of Origin** column:

MOTHER/FATHER – *de novo*: for each SNP probe locus overlapping an event, genotypes from the proband are assessed for their Mendelian inheritance against the parental genotypes. A probability score for each SNP position is then issued. An overall statistical metric, **Likelihood Ratio,** is calculated based on the array of probability scores; indicating how many times the event is more likely to be inherited from one parent versus the other. A minimum threshold of 10 times (10X) has been set for the **Likelihood Ratio**.

INSUFFICIENT_DATA - *de novo*: If the likelihood ratio is between 0.1 and 10.

NOTE: An empty cell indicates that an event overlaps an event in a parent.

Parent of Origin events for inherited CNV/AOH is based on a similarity score of proband and parent events. If one parent has a similar event, the **Parent of Origin** will be noted as either **MOTHER - Inherited** or **FATHER - Inherited**. If similar events exist in both parents, **Parent of Origin** will be marked as **CONFLICTING_DATA – Inherited**, indicating that one cannot be sure which parent it was inherited from as some probes can indicate **FATHER** and other probes, **MOTHER**.

A similar concept is used for inherited SeqVar events (whether the variant is present in one or both parents):

MOTHER – Inherited: Variant present in mother.

FATHER – Inherited: Variant present in father.

EITHER – Inherited: Variant present in both parents and is heterozygous.

BOTH – Inherited: Variant present in both parents and is homozygous.

Informative probes may be color coded in the BAF track based on the parent of origin, blue for paternal and pink for maternal, as indicated in **Figure 155**.

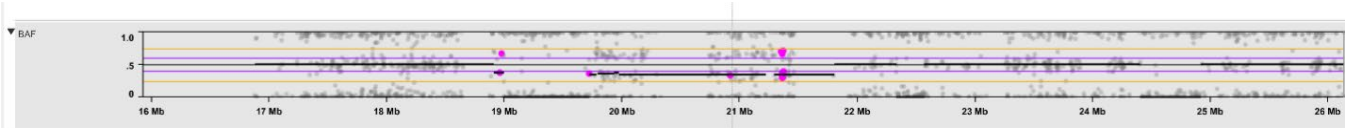


Figure 155. Maternally derived probes depicted as larger pink probes.

Track Display For Linked Samples

Once all relevant samples have been linked, click on the **Tracks** tab of the **Sample Review Preferences** window, and select the desired tracks to see in the display in the Linked Sample Relationship group, as shown in **Figure 156**.

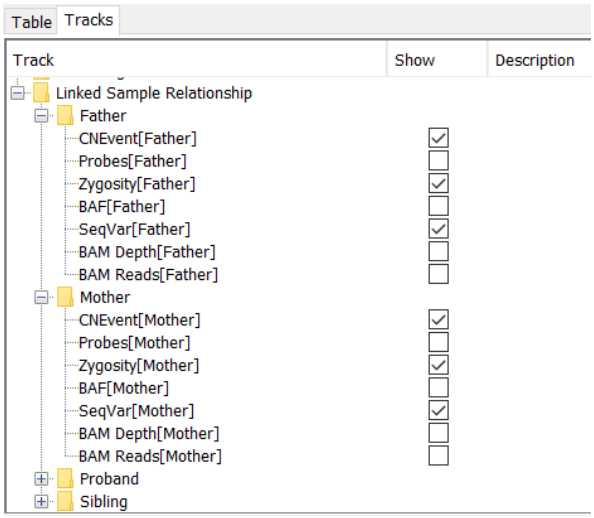


Figure 156. Linked sample relationship group.

Click and drag the tracks in the **Track Layout** section to change the display order of the tracks, as seen in **Figure 157**.

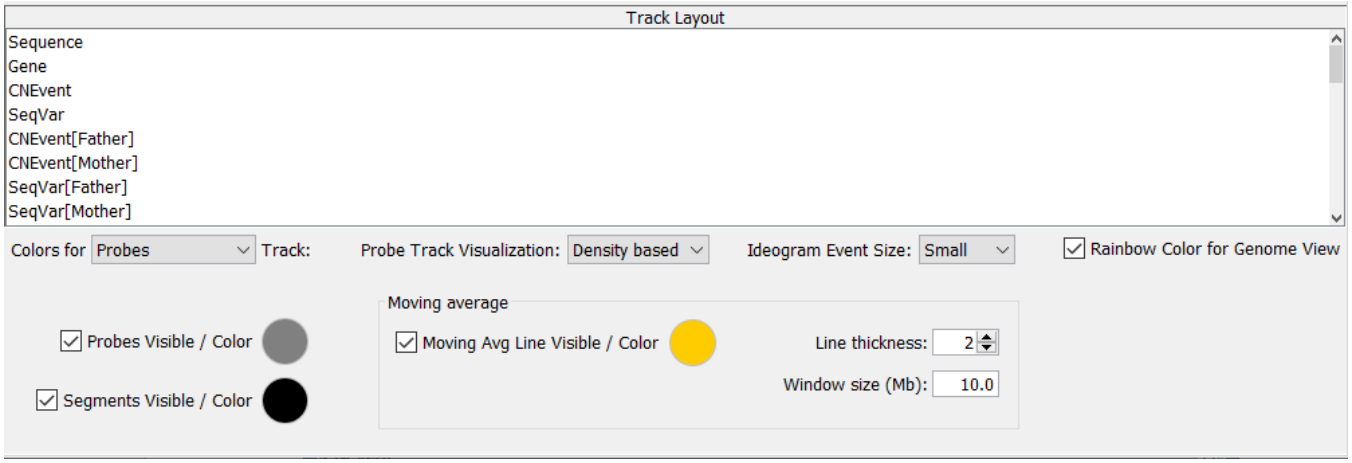


Figure 157. Track Layout image.

Now the available linked samples will be visible in the browser, as seen in **Figure 158**.

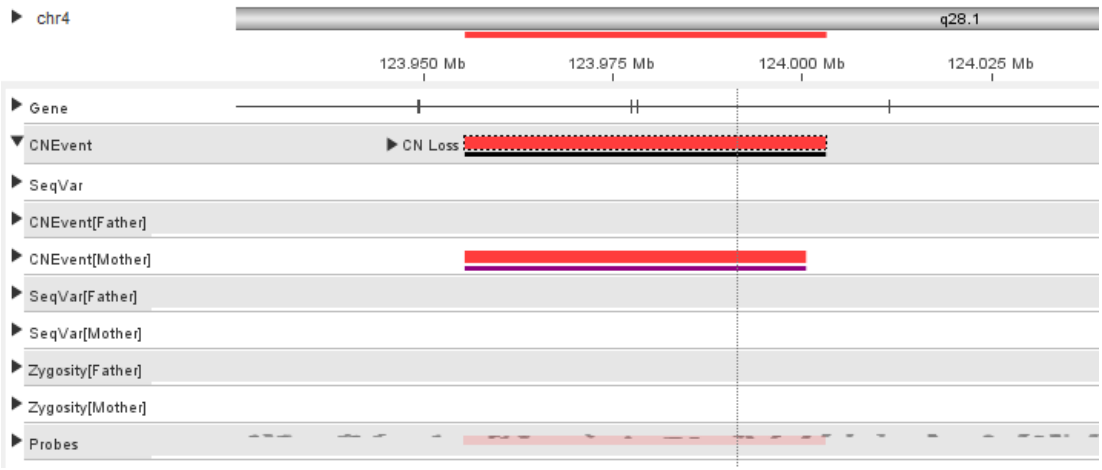


Figure 158. Available linked samples.

Detection of Uniparental Disomy (UPD) Events in Trios or Duos

For SNP arrays and WES/WGS (high coverage); can only be used as Duo or Trio sets with at least one parent sample linked to the Proband sample (not intended for oligo arrays, NGS panels or low-pass WGS).

NOTE: Significant improvements were made to UPD detection and parent of origin calculation in VIA 6.1 Build 14418, so it is recommended to use this build or higher.

Uniparental disomy occurs when an individual receives two copies of a chromosome, or a part of a chromosome, from one parent, and no copy from the other parent. UPD can occur via heterodisomy (hUPD), wherein the individual inherits a pair of non-identical chromosomes from a single parent, or isodisomy (isoUPD), in which a single chromatid from one parent is inherited and subsequently duplicated. Detection of UPD is of clinical relevance in the context of imprinting disorders and in the manifestation of recessive disorders where only one parent is a carrier.

Setting up Sample Relationships: For a Duo or Trio, sample relationships can be created by using a single Unique Linked Sample ID for all members of the family set and assigned the relevant Linked Sample Relationship for each member (e.g., Proband, Mother). This can be done by either editing sample attributes after uploading the samples through the **Home** page (**Sample Information** window) or by including these details in the batch upload file at the time of sample upload. Users can then run the UPD tool. UPD detection can be performed on a Proband with at least one parent in Edit mode using the **Check for UPD** button in the toolbar, as shown in **Figure 159**.

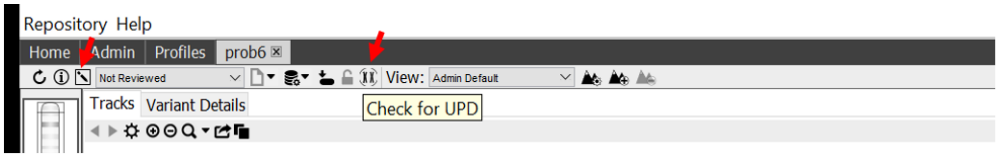


Figure 159. Check for **the UPD** button in Sample Review.

- The **UPD** button is active only in **Edit** mode.

- Upon clicking the **Check for UPD** button, existing AOH or AI calls in UPD regions will be deleted and replaced with hUPD or isoUPD calls; the AOH/AI calls will be moved to the **Deleted Events** table. A warning to this effect is visible as in **Figure 160**.

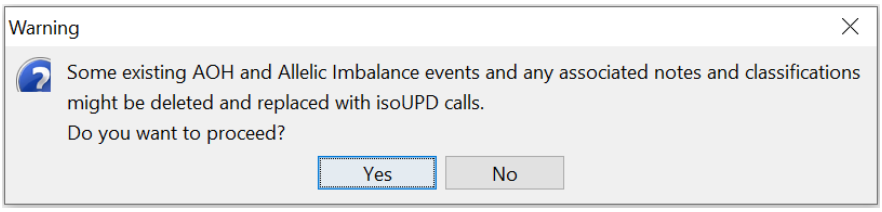


Figure 160. Deleted Events warning.

- UPD calling distinguishes UPD from autozygosity (homozygosity with two alleles identical by descent); regions of autozygosity remain labeled as AOH.
- The **Check for UPD** button is not visible if the sample does not have SNP data.
- UPD calling can only be run once; the button is disabled after processing is complete; to re-run UPD calling, the sample must be reset and re-processed.

Analysis of results: In case no UPD events are detected, a message box will appear stating so. If UPD events are detected, they will be listed in a pop-up window after the tool is run, as seen in **Figure 161**.

NOTE: The minimum LOH length processing parameter also applies to **iso/hUPD** events so that events smaller than the specified length in the processing type are not displayed.

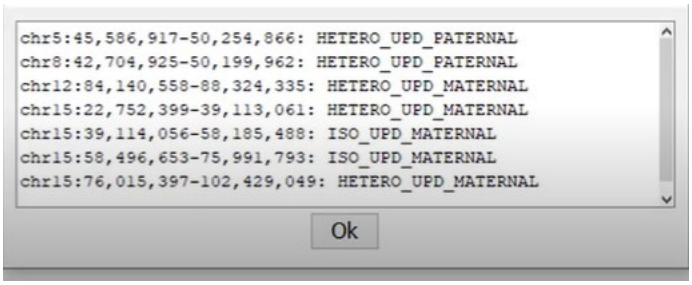


Figure 161. UPD events detected.

UPD events are represented as follows:

- Within the Ideogram view, a color scheme is used to denote UPD events as seen in **Figure 162** below.
  - **isoUPD:** Yellow painted chromosome with pink (maternal) or blue (paternal) edges representing informative homozygous SNPs.
  - **heteroUPD:** Yellow painted chromosome with a pink (maternal) or blue (paternal) central line representing informative heterozygous SNPs.

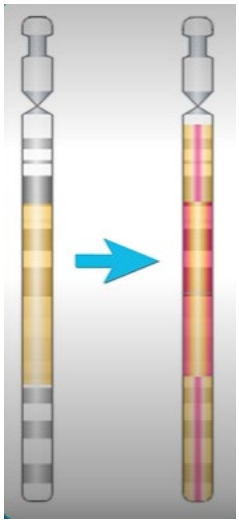


Figure 162. Ideogram color scheme for isoUPD and hUPD events.

- **Tracks view (see Figure 163):** While viewing a selected chromosome or region, parent of origin informative probes in the BAF track are displayed in the relevant color (pink=maternal; blue=paternal). A shade gradient is used, where washed out or lighter shaded probes indicate lower probability for the specified parental origin. Deep (dark) color probes indicate the highest evidence. Non-informative probes are colored grey. **IsoUPD** or **hUPD** labels are also visible in the Zygosity track using the same color coding as the ideogram view.

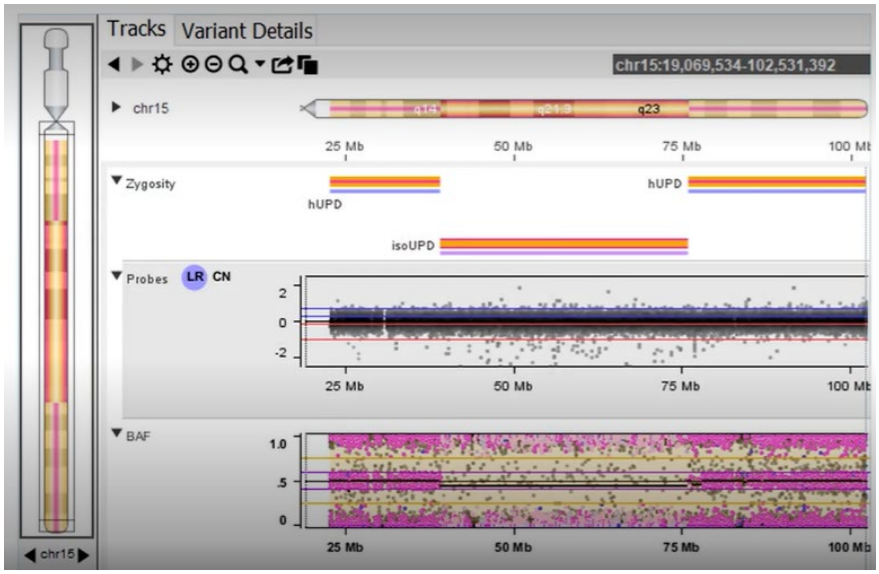


Figure 163. Tracks view.

- **Table view:** New calls, isoUPD, and hUPD are added in the **Event** field. The **Parent of Origin** column indicates the probability for the parent of origin as calculated. The exported sample review table file indicates if a UPD check was performed with the line, #UPD Processing Performed = true.
- **Filter pipeline:** The **Show/Hide Allelic Homozygosity Events** button now also applies to **hUPD** and **isoUPD** events. The Filtering pipeline (**Allelic Events** filter) allows displaying UPD events based on event type (Remove all hUPD/ isoUPD calls) or based on size/number of probes.

- **Variant Details tab:** In the **Variant Details** tab, the ISCN 2020 notation for the **UPD** event is provided, along with information on the **Parent of Origin** and the **Likelihood** ratio.

Creating Linked Samples

Linked Sample Relationships/Trios

A set of flexible attributes allows samples to be associated with each other and linked together in different ways.

TRIO/FAMILY RELATIONSHIPS

Samples may be linked together for Duo or Trio familial analyses with user-configured labels to specify the family relationships. A Trio linked sample relationship allows the sample to utilize specific features in VIA such as inheritance pattern filtering, parent of origin for events in the proband, differential coloring of probes inherited from mother vs. father, and the ability to view parental samples (e.g., events or probes) in the proband sample review.

Other Linked Sample Relationships

Samples may be linked together in other ways such as by patient or by samples (e.g., blood, primary tumor tissue, post treatment sample) taken from a single patient. This is often done with cancer samples where tissue is taken at different periods to track the cancer's progress. Another scenario using linked samples can be multiple embryos for preimplantation genetic testing (PGT). Linking samples allows for easier querying as well as the ability to visualize all samples in a single view.

Linked Sample Attributes

To associate samples with each other, use:

- **Linked Sample ID:** A unique ID for all samples belonging to that relationship; any alphanumeric string (can be a single word or multi-word)
- **Linked Sample Relationship:** Values defined by the Admin; Defaults in VIA: Proband, Mother, Father, Sibling (can be edited by the Admin), as seen in **Figure 164**.
- **Affected Status:** Clinical status of the patient, labeled as **Unaffected**, **Affected**, or **Unspecified** (unknown), as seen in **Figure 165**.

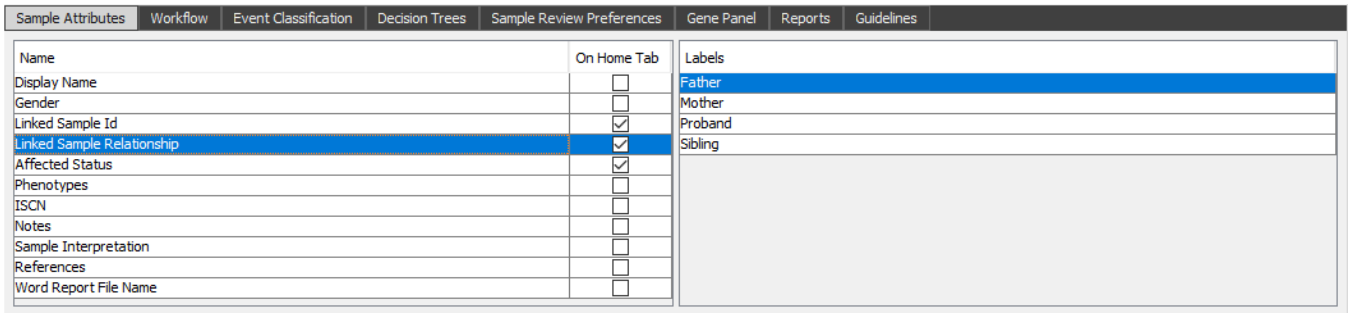


Figure 164. Default values for Linked Sample Relationship.

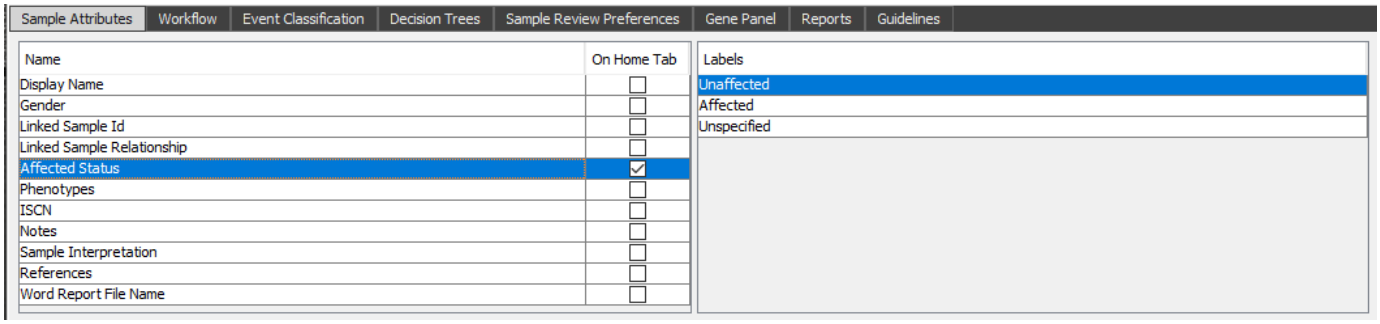


Figure 165. Default values for Affected Status.

Linked Samples Tab

The Admin can further customize the labels for Linked Samples to best represent lab workflow or local language. The **Linked Samples** Tab has a list of labels used for Linked Sample Relationship and the associated relationship meaning. Some family relationships need to have a standard meaning that the software can understand for calculations such as trio quality check, parent of origin, and recessive inheritance filtering. The relationship meaning values are selected using a dropdown and restricted to the following: Father, Mother, Proband, Sibling. For example, the Admin could create a new label called Dad which would have the meaning Father, shown in **Figure 166**.

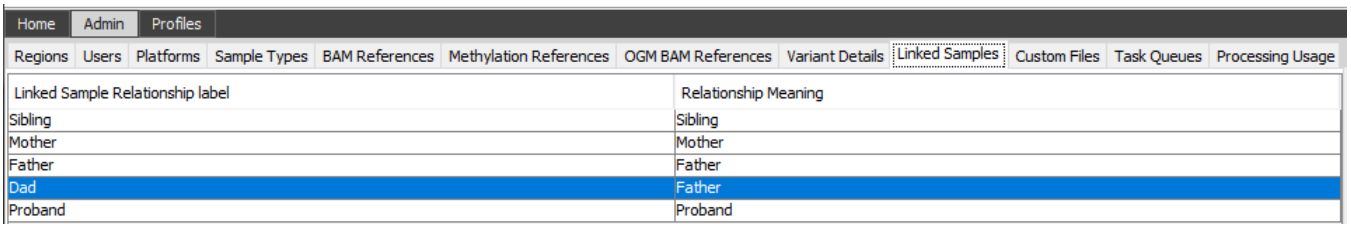


Figure 166. Relationship Meaning values.

To add a new label, click the **+** button and enter a label. Use the dropdown to select a relationship meaning and click **Save Changes**. Now the label Dad can be used in the **Sample Attributes** section. See **Figure 167** for an illustration of this attribute.

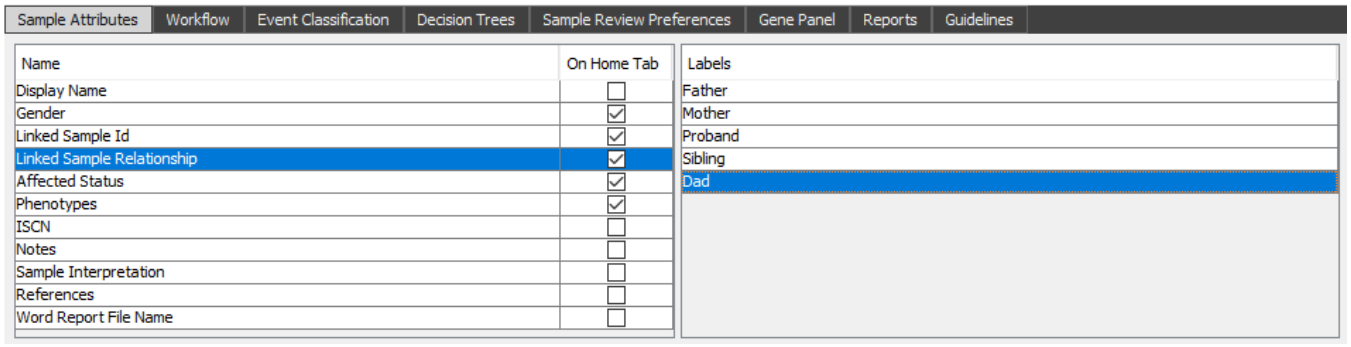


Figure 167. Adding a new label.

When any family calculations are performed on a sample, the software will know that Dad means Father and compute accordingly.

The Phenotypes Attribute

The Phenotypes attribute is a special attribute that allows association of HPO terms with the sample. This is specific to each individual sample and therefore added during or after sample upload. See the *Creating a Sample Type, Sample Loading and Processing* section for details on associating phenotypes.

Creating a Sample Type, Sample Loading and Processing

The Administrator assigns user privileges for loading and processing samples and can be contacted to modify user accounts if loading and processing are not available.

File Type Requirements and Data Modalities

A **Sample Class** defines the type of input data, or rather, the technology or platform from which the data is coming. The **Sample Classes** available are dependent on the VIA license which indicates the **Sample Classes** supported for a specific installation. Available Sample Classes include:

- **Array Only:** Input files are only composed of array data.
- **NGS and Array:** Input files may be array data and/or NGS data (BAM, VCF, JSON).
- **GxA-Cyto:** Input files must be GTC files from Illumina GSA-Cyto or GDA-Cyto arrays; GxA-Cyto final report files and standard GSA/GDA arrays cannot be processed with this class.
- **Low-Res WGS:** Input files are only composed of NGS data (BAM, VCF, JSON).
- **Methylation:** Input files are only composed of methylation data (IDAT).
- **OGM and NGS:** Input files may be OGM data and/or NGS data (BAM, VCF, JSON).

A sample type is defined by the **Sample Class**, **Genome Build**, and **Modality** (**CNV, SeqVar** and/or **SV**). When creating a sample type, a sample class, which cannot be changed later, must be selected.

- For CNV and AOH analysis from microarray data, use **Array Only** for sample class.
- For sequence variant analysis, use **NGS** and **Array**.
- For estimation of CNV from NGS, use **NGS** and **Array**.
- For structural variant analysis from OGM, use **OGM** and **NGS**.

OGM Sample Type

COPY NUMBER VARIANTS FOR OGM SAMPLES

There are three data type options for CNV from OGM data:

1. OGM VCF
2. OGM BAM Multiscale
3. OGM BAM Self-Reference

The **OGM VCF** data type, which includes copy number variants, is generated from the Solve algorithm and imported from Access. **OGM VCF** data files are saved under the *.ogm.vcf file format.

The **OGM BAM Multiscale** data type requires an **OGM BAM** reference which is created with a set of cytogenetically normal samples through the **BAM Multiscale Reference Builder**. **OGM BAM Multiscale** files are characterized with the ogm.bam file extension. As optional, indexed OGM BAM file content (`*.ogm.bam.bai`) will be automatically uploaded into VIA, together with the `ogm.bam` files, if both file formats are located with the same path.

The **OGM BAM Self-Reference** data type uses the proprietary self-reference algorithm to estimate copy number from OGM data. The input file for the **OGM BAM Self-Reference** data type is the `ogm.bam` file. As optional, indexed OGM BAM file content (`*.ogm.bam.bai`)will be automatically uploaded into VIA, together with the `ogm.bam` files, if both file formats are located with the same path.

CNV Platform for OGM Sample Type

OGM VCF PROCESSING TYPE

A mock-up of the OGM VCF Processing Type, Example **OGM Thresholds**, is installed by default in VIA. From this template, a functional copy can be created, edited, and associated with an **OGM** sample type. **OGM VCF Processing** settings are defined as shown below:

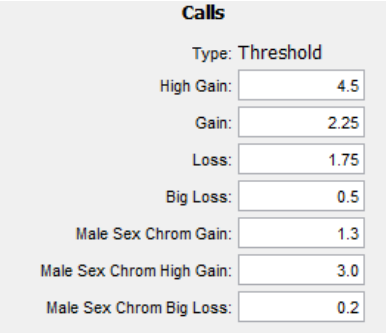


Figure 168. OGM VCF Processing settings.

OGM BAM MULTISCALE PROCESSING TYPE

A mock-up of the **OGM VCF Processing Type**, “Example OGM BAM Multiscale” is installed by default in VIA. From this template, a functional copy can be created, edited, and associated with an OGM sample type.

OGM BAM SELF-REFERENCE PROCESSING TYPE

A mock-up of the **OGM VCF Processing Type**, “Example OGM BAM Self-Reference” is installed by default in VIA. From this template, a functional copy can be created, edited, and associated with an OGM sample type.

BAF FROM OGM.BAM

For the **OGM BAM Multiscale** and **OGM BAM** Self-Reference, the probes on the **BAF** track are created with the parameters defined in the BAF from **OGM.BAM** processing setting. BAF from OGM.BAM has a minimum coverage threshold defined as **Reject labels with coverage less than**. Also, there is the ability to define a minimum MAPQ threshold under **Reject reads with MAPQ less than** setting. **The OGM cluster file for BAF** provides the location of the **SNP** probes for the **BAF** track.

Structural Variants for OGM Sample Type

OGM VCF is the only data type for structural variants (SV). The input file for OGM VCF is the `*.ogm.vcf` file.

Sequence Variants for OGM and NGS Sample Class

Along with CNV and SV, sequence variants can be uploaded in the same sample for sample types created under the OGM and NGS sample class.

SAMPLE ATTRIBUTES FOR OGM SAMPLE TYPE

The sample attributes below are available by default in the OGM sample type.

Table 12. Sample attributes.

Attribute Name	Description
Job ID	The Job number in Access
OGM Reference	The reference (.cmap) file used
Access version	Version number of Access software that processed the sample files
Solve version	Version number of Solve software that processed the sample files
Job type	The job type specified in Access (such as Guided Assembly, De Novo Assembly)
N50 (>=150kbp)	The molecule length N50 for all molecules that are ≥ 150 kbp in length.
Total length (>=150kbp)	The total amount of DNA that is detected in this flowcell across all runs of this chip
Map rate	The percentage of molecules that map to the reference for molecules ≥ 150 kbp. If no reference genome is provided, the metric is blank.
Average Label Density (>=150kbp)	The number of labels that are detected by the image detection algorithm per 100 kbp of DNA length for molecules ≥ 150 kbp.
Effective coverage of reference	

Although the attributes are available by default for the OGM sample type, the fields must be selected in Access to import into VIA.

Uploading and Processing OGM samples

OGM samples can be uploaded directly from Access by selecting the option to **Upload to VIA** when submitting a sample for processing. **NOTE**: VIA server settings should be set up in the **System Services Settings** in Access to connect to VIA Server. Details on how to set up the VIA connect can be found in the *Bionano Access User Guide*. Once the sample is uploaded, then the sample can be processed in VIA by searching for the uploaded sample(s) and selecting it(them) to process.

Alternatively, OGM samples can be manually uploaded into VIA through the Data or the Batch Import method. The *.ogm.bam (with the accompanying *.ogm.bam.bai) and *.ogm.vcf file for each sample are required to manually upload into VIA. These samples files can be downloaded from Access by selecting the sample in **Home > Analysis > Project Name** and navigating to the **Options** section.

Samples > Upload > Data Method

If OGM BAM Multiscale or **OGM BAM Self-Reference** is selected as the data type for CNV and **OGM VCF** is selected as the data type for SV, then the prefix name of the ogm.bam and the prefix name for the paired `ogm.bam.bai` and `ogm.vcf` files must match for the files to be loaded as one sample. For example, if uploading a file named sample1.ogm.bam, the prefix name “sample1” should be used for the paired files such that the names of the files would in this example case be sample1.ogm.bam.bai and sample1.ogm.vcf. Matching the prefix names for the paired `ogm.bam, ogm.bam.bai` and `ogm.vcf` files is only necessary when uploading samples using the **Samples > Data** method. When using the batch import method, the files specified in each row will load as one sample.

Single Sample Loading

CNV specifies the assay/platform/manufacturer of data for this sample type. Data can be from NGS using the BAM MSR or Self-Reference algorithms or it can be an array platform. Dropdown fields (**Data Type**, **Manufacturer**, **Assay Name**) allow one to specify details of this component. In addition, a processing type needs to be selected using the **Processing Type** field for this modality. Multiple processing settings may be associated with a single **Sample Type CNV** modality, and the user chooses which one to apply during processing.

SeqVar specifies the type of input data for sequence variants for this sample type. Input data can be unannotated VCF, annotated VCF, annotated JSON, or an unannotated file that is to be annotated using Nirvana linked to VIA. In addition, a processing type needs to be selected via the **Processing Type** field for this modality. Multiple processing settings may be associated with a single **Sample Type SeqVar** modality, and the user chooses which one to apply during processing.

SV specifies the input data is coming from the OGM. OGM VCF is the only Data Type for structural variants (SV). The input file for OGM VCF is the `*.ogm.vcf` file.

The processing type specifies the parameters to be used for processing the CNV or SeqVar component of a sample and each modality (CNV/SeqVar) of a sample type must have at least one associated processing type. The parameters are specified in the **Platforms** tab, where the user selects which processing type to apply during sample loading and processing.

The requirements for loading and processing data of the different classes are based on the type of license purchased. One can have a license to process samples belonging to one class only or to multiple classes. Once a sample type has been added by an administrator, data can be loaded into the VIA database through the main VIA interface (**Home** tab). Only users with permission to load data can perform this task (user permissions are set by the VIA Administrator). Data loading is performed using the **Samples** dropdown at the top right of the window. The **Upload** button will be grayed out for those without the correct privileges.

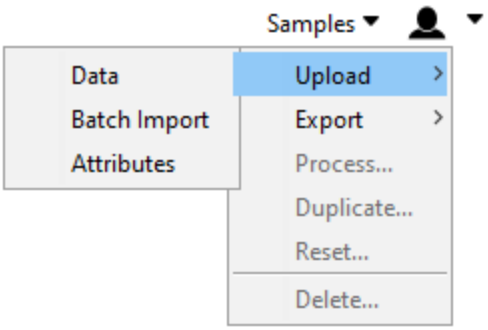


Figure 169. Samples dropdown menu.

There are two ways to load data. The simplest method is to load raw data by selecting the input files using a file chooser. In some cases, data upload with a sample descriptor (a text file containing names of raw data files to load) may be preferred, for example when uploading large volumes of legacy data or when a user needs to load samples and attributes at the same time.

To load data of different sample types (e.g., Affymetrix CytoScan arrays, Illumina CytoSNP 12 arrays) using the file chooser, all files of one sample type must be loaded before data of another sample type can be loaded, e.g., if there are 10 Illumina CytoSNP 12 files and 15 Affymetrix CytoScan HD files to load then first select Illumina CytoSNP 12 in the **Sample Type** dropdown and add the 10 Illumina final report files and click **Upload**. Next, again select **Samples > Upload > Data…** and then select the appropriate sample type. Select the 15 Affymetrix CytoScan HD sample files to upload.

NGS data on AWS stored on Amazon S3 can be accessed directly by VIA clients and a processing server.

Batch Loading

The batch import feature will support files (BAM, VCF or JSON) located on s3. To specify files, use the following syntax.

NOTE: The path is case-sensitive; s3 must be in lower case, as shown below:

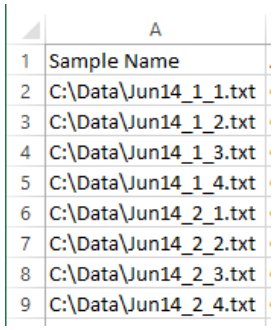
```
s3://s3-bucket-name/filename.VCF

s3://s3-bucket-name/filename.bam
```

To quickly load many samples in batch rather than selecting one by one, users will need to create a tab delimited text file (descriptor) that contains the file names and sample names (see **Figure 172** below). There are two different tools available for batch upload using a text file based on the types of samples that are being uploaded and the processing settings to be applied:

1. All samples are of the same sample type (e.g., all are Affymetrix OncoScan samples) and the same sample processing setting is to be applied to all the samples.
2. Samples are a mixture of different sample types (e.g., Affymetrix OncoScan, Illumina Infinium CytoSNP 850k – Postnatal, Illumina Infinium CytoSNP 850k – Prenatal) and/or different processing settings are to be applied to different samples. Some of the OncoScan samples should be processed with the Affymetrix

TuScan algorithm and others should be processed with the SNP-FASST2 algorithm. This method also allows loading of multiple modalities together (array data, VCF/Nirvana JSON file, BAM file).



	A
1	Sample Name
2	C:\Data\Jun14_1_1.txt
3	C:\Data\Jun14_1_2.txt
4	C:\Data\Jun14_1_3.txt
5	C:\Data\Jun14_1_4.txt
6	C:\Data\Jun14_2_1.txt
7	C:\Data\Jun14_2_2.txt
8	C:\Data\Jun14_2_3.txt
9	C:\Data\Jun14_2_4.txt

Figure 170. File names and sample names.

NOTE: For sample names, certain characters, such as the colon, forward slash, back slash, and others, cannot be used in the sample name. The software validates sample names before loading. If the sample name contains the restricted characters, an error message is displayed, and the sample will not be loaded.

Batch Uploading Samples of the Same Sample Type

To upload in batch, samples of the same sample type and same processing settings, a sample descriptor file is needed. This is a tab-delimited text file containing sample names, file locations, and optional attributes.

The Sample Descriptor Format requires a column header row. **Table 13** below describes the required and optional columns for a sample descriptor.

Table 13. Required and optional columns for a sample descriptor.

Column Header	Description
Sample Name	The first column must contain sample names. This can be just the file name itself (if the sample descriptor is in the same directory) or the full file path. This column can be called Sample Name and the values must be unique names. Two samples cannot have the same name.Note for Illumina data, the sample name in the first column must match values in the Sample Name or Sample ID column of the final report file. If it does not, an error message will be displayed after the files are parsed. Another column must have the header Filename. If the sample file is in the same directory as the descriptor file then just the filename can be listed. If the file is in any other directory, the full file path must be used in this column. ***Required***
Reference File	Specifies the name of the reference file to use. To ensure the correct reference name is used, first use the **Upload->Data** tool to view all available reference files and enter the relevant reference name into the descriptor. If the appropriate reference files are not displayed here, please contact your VIA Administrator. ***Required for Agilent CGH+SNP arrays and for sample types deriving CNV from BAM files (Data Type = BAM Multiscale)***
Filename	Location of the sample file for loading CNV data. Just the file name or full file path can be used (e.g., C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_01.OSCHP). ***Required for Illumina Final Report files***

Display Name	The sample display name – the name to be displayed for this sample within VIA. Typically specified manually in the **Sample Info** window. ***Optional***
Panel	Enter the name of a gene panel to be pre-selected during sample review. The gene panel must already be associated with a sample type prior to loading the descriptor. ***Optional***
[sample attribute columns]	Column headers are the sample attribute names. Values in the columns would be the sample attribute labels. Note that the attributes listed in the descriptor file must match those associated with a sample type in the VIA system. The attributes associated with a sample type can be found in the **Sample Types** tab, in the **Sample Attributes** subtab. Note that the Gender attribute can only have values Male, Female, and Unspecified. One cannot use M to designate Male, for example. ***Optional***

Agilent CGH+SNP arrays and BAM files for deriving CNV, a column called **Reference File** is required with the reference file specified here. **Figure 173** and **Figure 174** illustrate how to view all available reference files and then enter the relevant file name into the descriptor.

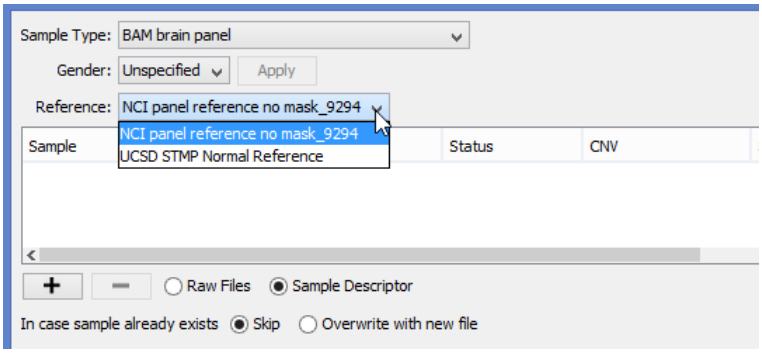


Figure 171. The **Upload Sample Data** window displays available Reference files for a sample type for CNV estimation from NGS.

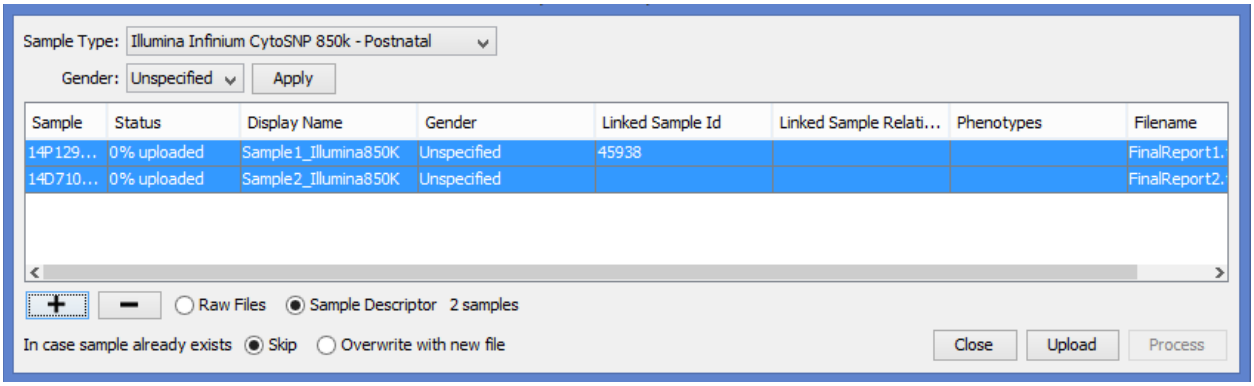


Figure 172. Upload window shows **Illumina Final Report** files to be uploaded using a Sample Descriptor.

	A	B	C	D	E	F
1	Sample Name	Filename	Display Name	Gender	Linked Sample	Linked Sample Relationship
2	14P1292CX_HE4	FinalReport1.t	Sample1_Illumina850K	Unspecified	45938	
3	14D7103A1_HE4	FinalReport2.t	Sample2_Illumina850K	Unspecified		

Figure 173. The corresponding sample descriptor file opened in Excel.

If sample attributes are to be uploaded at the same time, the sample attributes can be specified in the sample descriptor file. An example of a sample descriptor file with attributes Gender and Age is seen in **Figure 176**.

Sample Name	Display Name	Gender	Age
C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_01.OSCHP	Sample1	Unspecified	67
C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_02.OSCHP	Sample2	Unspecified	71
C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_03.OSCHP	Sample3	Unspecified	71
C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_04.OSCHP	Sample4	Unspecified	59
C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_05.OSCHP	Sample5	Unspecified	80
C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_06.OSCHP	Sample6	Unspecified	66

Figure 174. Gender and Age Attributes in a Descriptor File.

The attributes listed in the descriptor file must match those associated with a sample type in the VIA system.

Figure 177 below shows the attributes associated with the **Affymetrix OncoScan** sample type, the same sample type in the above descriptor file.

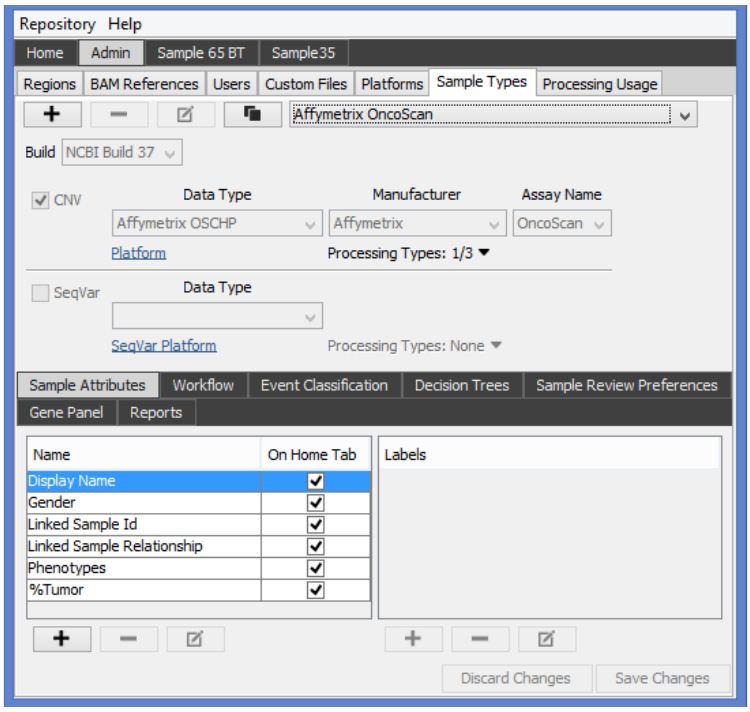


Figure 175. Attribute associated with the Affymetrix OncoScan sample type.

If any attributes listed in the descriptor file are not defined in VIA, an error message will display indicating unknown attributes were included in the file and the file will not be loaded. For example, if the descriptor has a column called Age (as in the above example), this file will not be loaded because Age is not an attribute associated with the **Affymetrix OncoScan** sample type. In this case, an error message, such as the one in **Figure 178**, will be displayed. If additional attributes associated with a sample type are needed but are not listed, contact the Administrator.

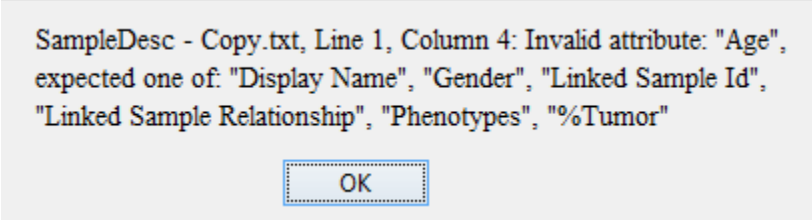


Figure 176. Error message.

LOADING THE SAMPLE DESCRIPTOR

To load data:

1. Select **Samples > Upload > Data…** from the **Samples** button.
2. In the **Upload Sample Data** window, first select the sample type from the dropdown at the top of the window, as shown in **Figure 179** below.

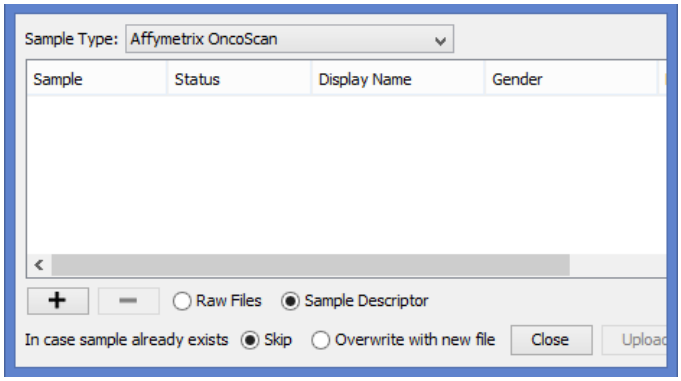


Figure 177. Loading the sample descriptor.

3. Ensure that **Sample Descriptor** is selected at the bottom and then click on the **Add Files** button (**+** button). This opens a file chooser where the descriptor file can be selected, as shown in **Figure 180**.

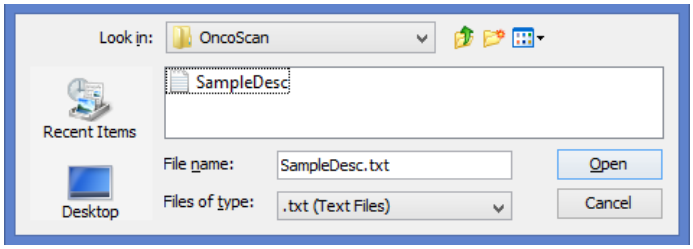


Figure 178. File chooser.

4. Once the descriptor file has been selected, the samples listed in the file will be displayed in the **Upload** window, as seen below in **Figure 181**.

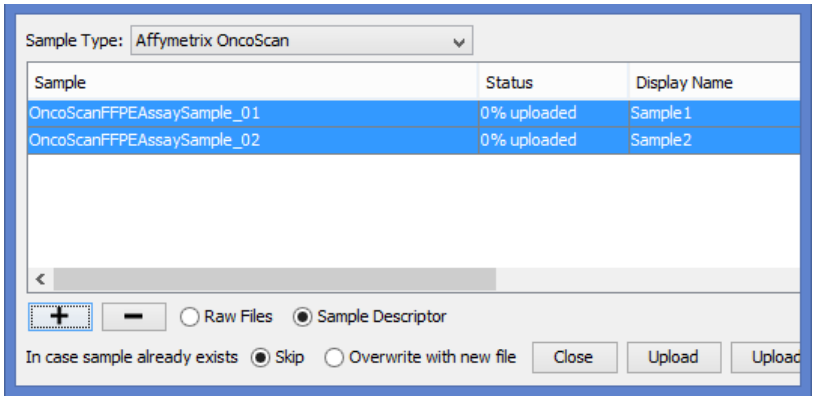


Figure 179. File display.

Click **Upload** to copy the sample files to the database, as shown in **Figure 182**. The **Status** field shows file upload progress (as percent uploaded) while the files are being uploaded from the local machine to the server, as shown in **Figure 183**. If additional attributes associated with a sample type are needed but are not listed, contact the Administrator. While data is being loaded, one can continue using VIA to browse other cases in the database.

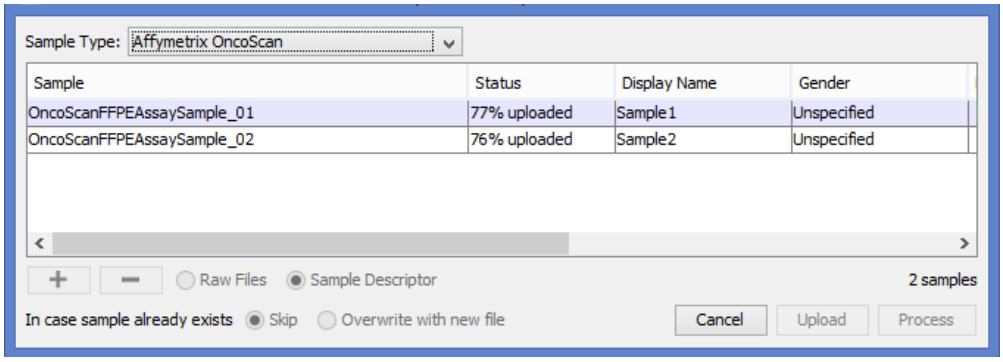


Figure 180. Uploading to copy sample files.

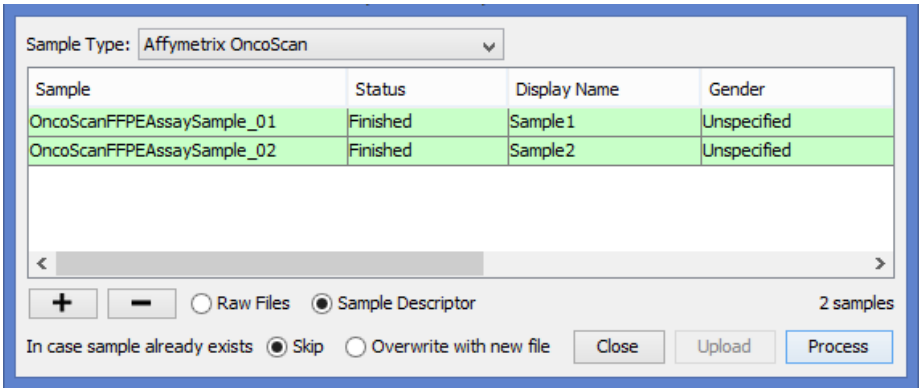


Figure 181. Status field.

Batch Import of Samples of Different Sample Types and/or Importing Multiple Modalities

If a user wanted to load samples of different types or multiple modalities in batch, a different tool should be used. Using the **Upload > Data** tool and then selecting **Sample Descriptor** will not work as this requires selection of a **Sample Type** from the **Upload** interface and all samples in the descriptor need to be of this sample type.

If, for example, there are OncoScan samples and Illumina samples that need to be loaded and one does not want to proceed a single sample type at a time, use the **Upload > Batch Import** tool available with the **Samples** button in the top right of the window, as seen in **Figure 184**. Choosing Batch Import opens a window from which a descriptor file should be selected, as shown in **Figure 185**.

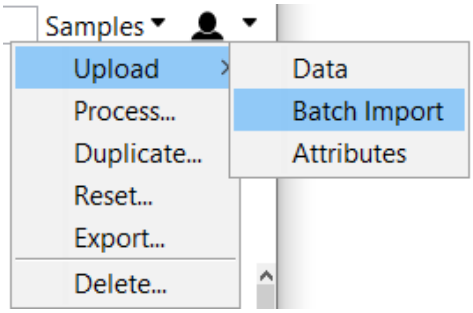


Figure 182. Batch Import.

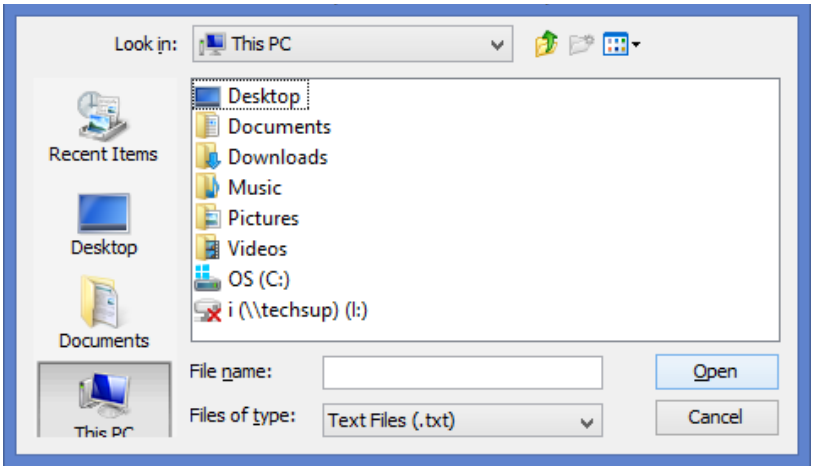


Figure 183. Selecting a prepared batch descriptor.

BATCH DESCRIPTOR FORMAT

The batch descriptor file is a tab delimited text file containing sample names, file locations, settings, attributes, sequence variant and BAM files. Templates for the descriptors are available in the VIA Client/Templates/Batch Import folder of the client installation directory. **Table 14** below describes the columns (required/optional) for the sample descriptor.

Table 14. Batch Descriptor Format.

Column Header	**Description**
Sample Name	Name of the sample. **NOTE:** For Illumina samples where multiple Illumina samples are in one Final Report file, the name listed here must match the names in the Sample Name column of the Final Report file. When the software encounters a name that does not exist in the final report file, an error is indicated in the status column of the upload window and no further samples from the Final Report file will be parsed. If even one sample name does not match, none of the samples from the Final Report file will be loaded. ***Required***
Filename	Location of the sample file for loading CNV data. Can be just the file name (if in the same folder as the descriptor) or full file path (e.g., C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_01.OSCHP). This is used for loading array data or BAM for deriving CNV. The Filename need not match the Sample Name nor the file name of seq var or BAM files (Seq Var File or BAM File columns, respectively). ***Required if loading CNV data and for Illumina Final Report files***
Sample Type	The VIA Sample Type of this sample. Please make sure the name matches exactly to an existing sample type in VIA. ***Required***
Processing Setting	A processing type for the indicated CNV modality of the sample type. Note that some samples have multiple processing types so make sure to use the correct one for each sample. ***Required if loading CNV data***
Reference	Specifies the name of the reference file to use. To ensure the correct reference name is used, first use the **Upload->Data** tool to view all available reference files and enter the relevant reference name into the descriptor. If the appropriate reference files are not displayed here, please contact your VIA Administrator. ***Required for Agilent CGH+SNP arrays and for sample types deriving CNV from BAM files (Data Type = BAM Multiscale)***
Control Sample	Specifies the name of the control sample for ImaGene Data Types. ***Required for ImaGene Data Type only***
Seq Var File	Location of the sequence variants file to associate with the sample. Just the file name or full file path can be used (e.g., C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_01.VCF). The file name here need not match the Sample Name nor the file name of the file for CNV estimation of BAM files (Filename or BAM File columns, respectively). ***Required when loading sequence variants***
Seq Var Setting	A processing type for the indicated SeqVar modality. Note that some samples have multiple processing types so make sure to use the correct one for each sample. ***Required when Seq Var File is specified for a sample***
BAM File	Location of the BAM file to associate with the sample. Just the file name or full file path can be used (e.g., C:\Projects and Data\Data\OncoScan\OncoScanFFPEAssaySample_01.BAM). This is only used when loading BAM files to view read depth (e.g., associating with an array or seq var file). This column is not used to load BAM files for deriving CNV. The file name here need not match the Sample Name nor the file name of the file for CNV estimation file or BAM files (Seq Var File or Filename columns, respectively). ***Optional***
Panel	Enter the name of a gene panel to be pre-selected during sample review. The gene panel must already be associated with a sample type prior to loading the descriptor. ***Optional***
[sample attribute columns]	Column headers are the sample attribute names. Values in the columns would be the sample attribute labels. Note that the attributes listed in the descriptor file must match those associated with a sample type in the VIA system. The attributes associated with a sample type can be found in the **Sample Types** tab, in the **Sample Attributes** subtab. ***Optional***

An example descriptor file (tab delimited txt file) opened in Excel so that the different fields and data are clearly visible is seen in **Figure 186.**

Sample Name	Filename	Sample Type	Processing Setting	Gender
14P1292CX_HE42	C:\Users\sverma\Do	Illumina Infinium CytoSNP 850k - Postnatal	Illumina Infinium CytoSNP 850k - Postnatal	Unspecified
Sample1	C:\Projects and Data	Affymetrix OncoScan	Affymetrix Oncoscan	Unspecified
Sample2	C:\Projects and Data	Affymetrix OncoScan	Affymetrix Oncoscan	Unspecified

Figure 184. Example Descriptor file.

An example descriptor file containing **BAM** and **SeqVar** columns is seen in **Figure 187**. Note that some samples can be missing the BAM file or BAM and SeqVar; in such cases only CNV data will be loaded.

Sample Name	Filename	Sample Type	Processing Setting	Gender	Seq Var File	Seq Var Setting	BAM File
Sample1	C:\Projects and Data	Illumina Infiniun	Illumina Infinium CytoS	Male	SV_1	VCF processing	C:\Projects and Da
Sample2	C:\Projects and Data	Illumina Infiniun	Illumina Infinium CytoS	Female	SV_2	VCF processing	
Sample3	C:\Projects and Data	Illumina Infiniun	Illumina Infinium CytoS	Female	SV_3		

Figure 185. File containing BAM and SeqVar columns.

Seq Var only or Seq Var and BAM files can be loaded without associated CNV data. Note that some samples have no value for Filename; in such cases the column will be ignored and only the vcf or vcf+BAM files will be loaded. An example descriptor file showing that Seq Var only (sample4) or Seq Var and BAM (sample5) files can be loaded without CNV data (**Filename** column is empty) is shown in **Figure 188**.

Sample Name	Filename	Sample Type	Processing Setting	Gender	Seq Var File	Seq Var Setting	BAM File
sample4		Illumina_B		Male	sample4.vcf	VCF3	C:\Downloads\Cyt
sample5		Illumina_B	Illumina 850k Cance	Female	sample5.vcf	VCF3	
sample6	C:\Downloads	Illumina_B	Illumina 850k Cance	Female			

Figure 186. Files loaded without CNV data.

LOADING THE BATCH DESCRIPTOR

The software parses the descriptor and then loads all non-Illumina samples; it then proceeds to load the Illumina samples.

To Load data:

1. Select **Samples > Upload > Batch Import** from the **Samples** button at the top right of the window.

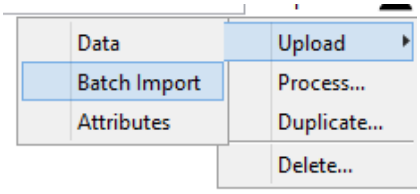


Figure 187. Batch Import navigation menu.

2. In the **Select Prepared Batch Descriptor** window, navigate to the folder containing the descriptor, select it and click **Open**.

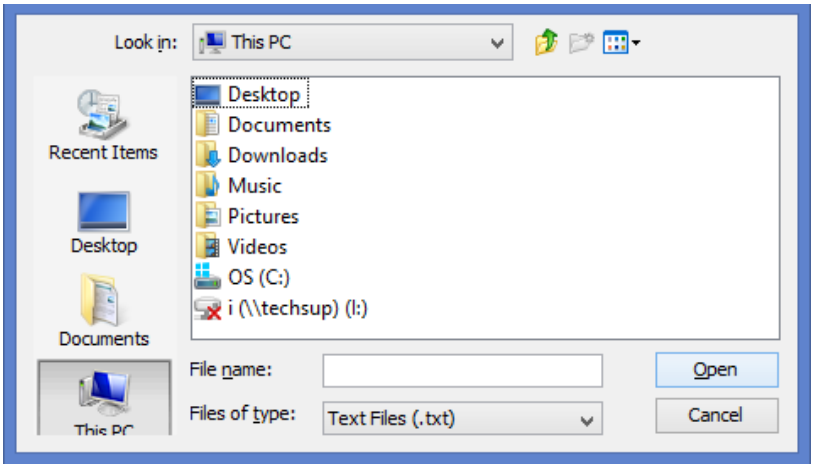


Figure 188. Folder containing the descriptor.

3. The **Batch Import** window will display contents of the descriptor file.

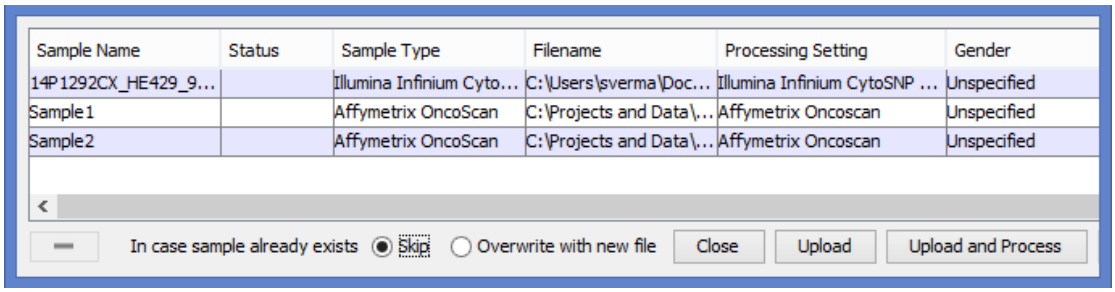


Figure 189. Batch Import window.

4. If any samples should not be uploaded, highlight the row, and click the minus (**-**) button to remove the sample(s) from the **Import** list.

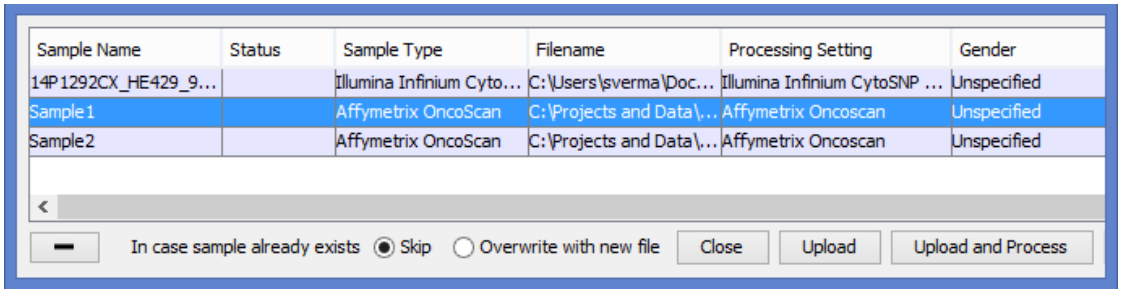


Figure 190. Import list.

5. To load the samples only, click the **Upload** button. To immediately process samples after upload is complete, click **Upload and Process**.

   During the loading process, the **Status** column will indicate which state each sample is in.

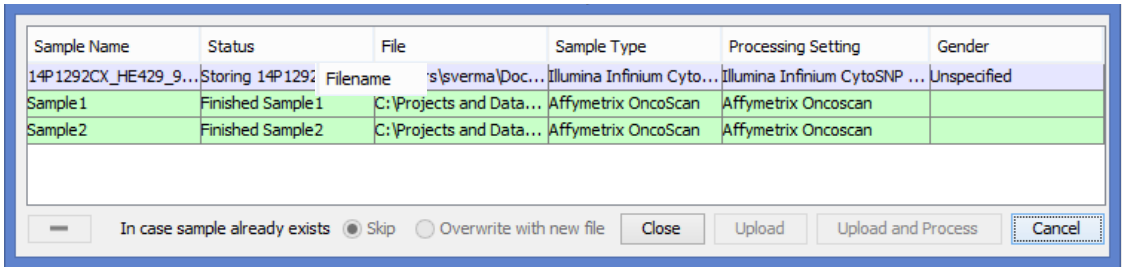


Figure 191. Sample upload status.

If this window is closed during the upload and/or processing steps, loading/processing will continue in the background and the status of the import can be reviewed by opening the window again from the **Samples->Upload** menu:

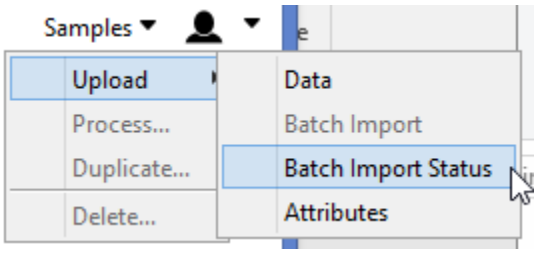


Figure 192. Samples Upload menu

If the **Batch Import Status** option is not available in the menu, this means that upload is finished.

6. If a user needs to cancel the upload, click the **Cancel** button. The samples not yet uploaded will display **Cancelled** in the **Status** column:

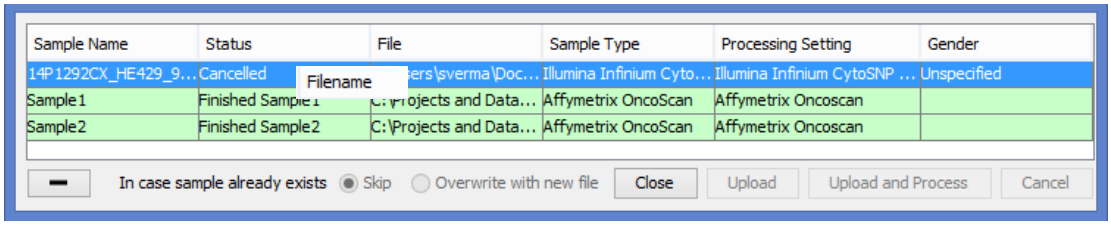


Figure 193. Cancelled upload status.

7. When loading is finished, the **Batch Import** window will display **Finished** in the **Status** column of each sample:

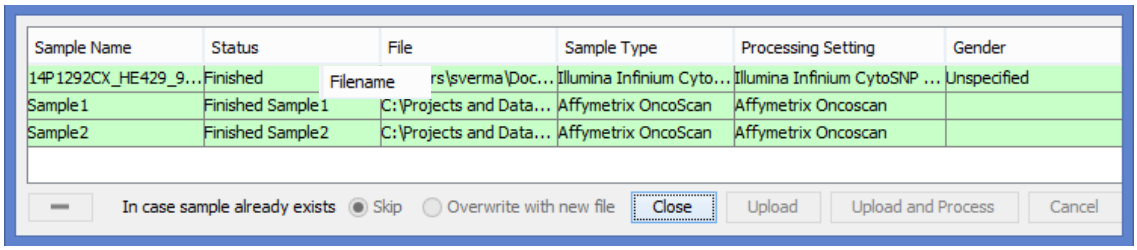


Figure 194. Finished upload status.

8. Click **Close** to close the window. If samples still need to be processed, go to **Samples->Process** to process the samples.

SPECIAL CASE: ASSOCIATING A BAM FILE TO AN EXISTING SAMPLE

If an array sample of type CNV and SeqVar is already in VIA and a BAM file needs to be added to it using batch import, the **Upload and Process** button must be used to load and process the BAM file. The **Upload** button cannot be used in this case as the read depth must be calculated for the file to upload. Read depth calculation occurs in the processing component so if the **Upload** button is used, the BAM file will not get associated with the existing sample.

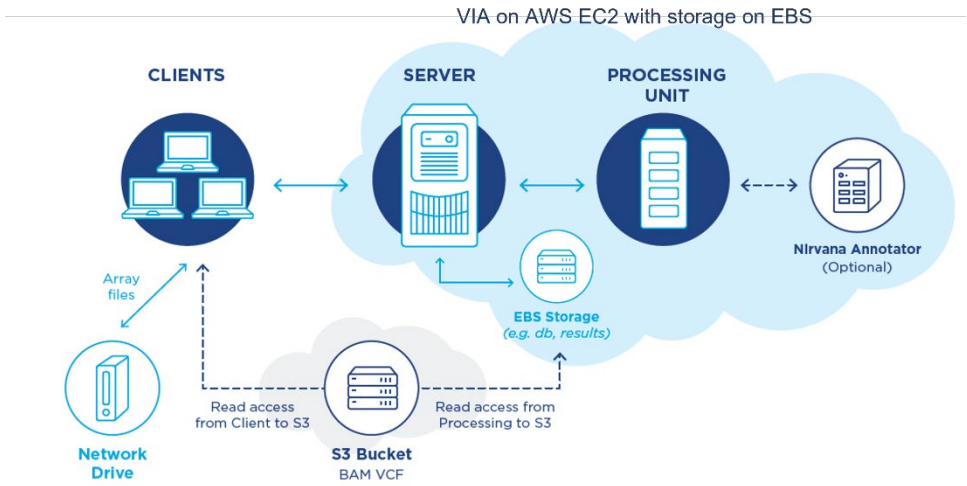


Figure 195. BAM files.

To allow such features, many steps need to be taken to set up the system and S3 appropriately for all components of VIA to have access to each other. S3 bucket credentials can be supplied to the VIA server in several diverse ways (see `https://docs.aws.amazon.com/sdk-for-java/v1/developer-guide/credentials.html`). Contact FAS to get help in setting up such a system.

LOADING FILES FOR CNV CALLING USING THE BAM MULTISCALE REFERENCE METHOD

BAM files can be loaded to derive copy number variants using CNV segmentation and calling algorithms. Before a file can be loaded, it must have a reference file associated with it. This reference file is created using the Multiscale BAM Reference Builder application installed separately (see the *Generating BAF values from BAM files* section). Once a reference file is generated, loaded into VIA, and associated with a sample type, the NGS sample can be loaded into VIA. VCF or JSON files for sequence variants to be associated with the BAM file can be loaded at the same time or later.

On the **Home** tab, select **Samples > Upload > Data**. In the **Upload Data** window, select the appropriate sample type and then the reference file for this sample, as shown in **Figure 198**. If the user does not see an appropriate reference file, the VIA Administrator should be contacted.

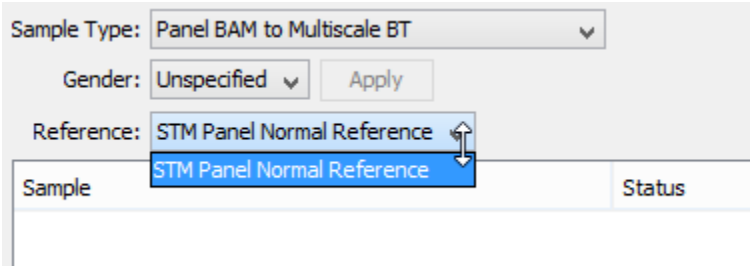


Figure 196. Sample type and reference file.

Click on the **+** button to select files to load using the **File Chooser**. Note that the file types displayed are only those relevant to BAM files, as shown in **Figure 199**, and that .bai files are not displayed for selection but need to be present in the same folder.

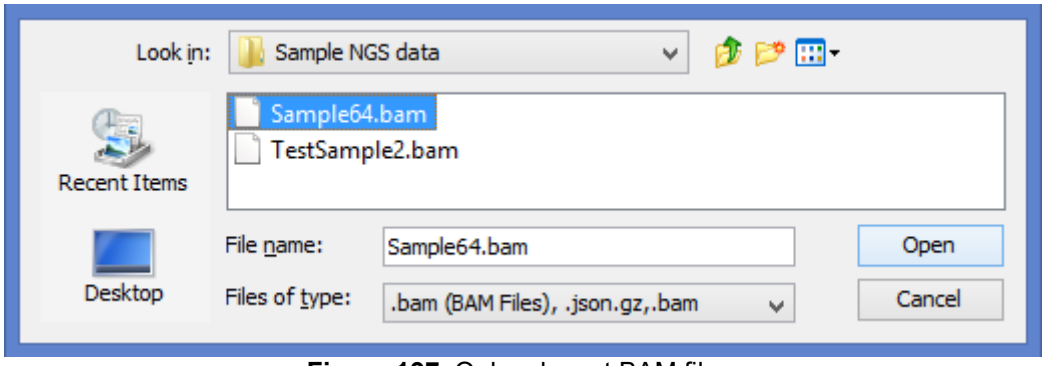


Figure 197. Only relevant BAM files.

Once files are selected, they will appear in the list of files to be loaded with the boxes marked off for the different modalities based on the sample type selected and files selected for upload. In **Figure 200**, **SeqVar** files were also selected in addition to BAM files.

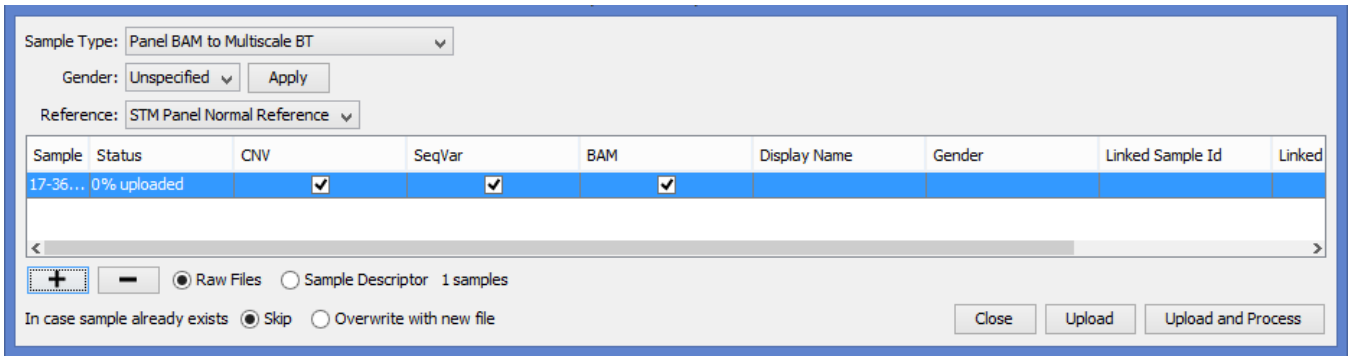


Figure 198. SeqVar and BAM files.

If **Upload and Process** is selected, processing will begin immediately after upload is finished. Hovering over this button will show a tool tip indicating the **Processing** settings that will be used for the CNV and the SeqVar processing, seen in **Figure 201**.

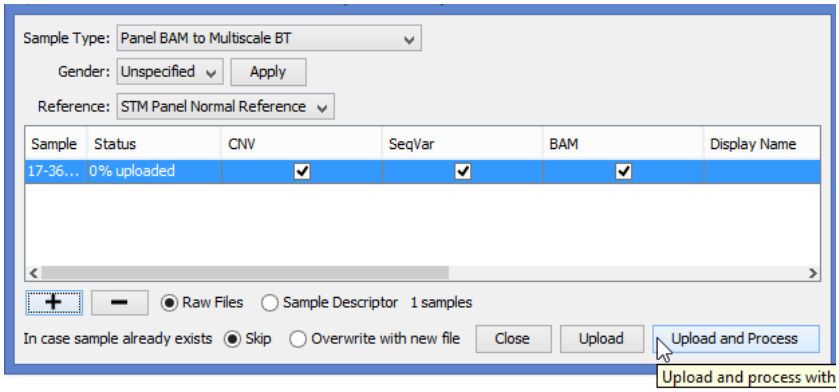


Figure 199. Tool tip.

If the **Upload** rather than **Upload and Process** button is selected, the files will upload but not process and the **Process** button will become active when the files have finished uploading, as shown in **Figure 202**.

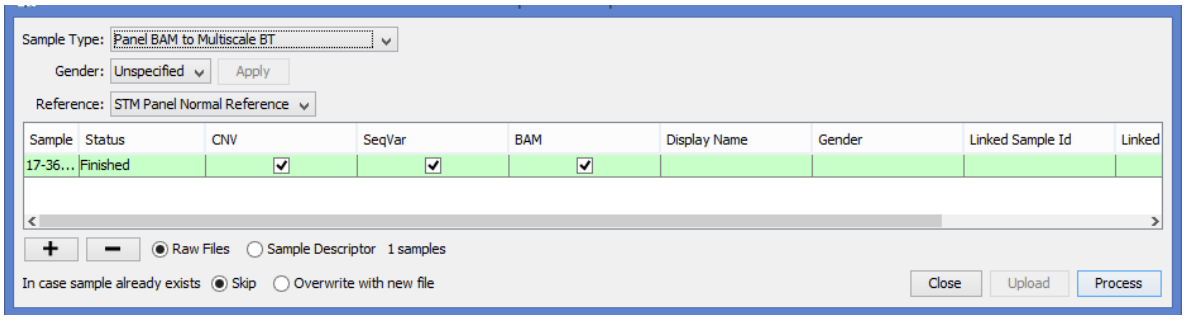


Figure 200. Selecting only the **Upload** button.

NOTE: BAM files are not actually physically loaded to the database; instead, the location of the file on the drive is stored in the database. If this file is moved, reads will not be displayed and the VIA system will prompt the user to provide a new location for the physical files, re-associating them with the sample.

Clicking on **Process** will bring up the **Processing** window where various settings need to be selected, as seen in **Figure 203**.

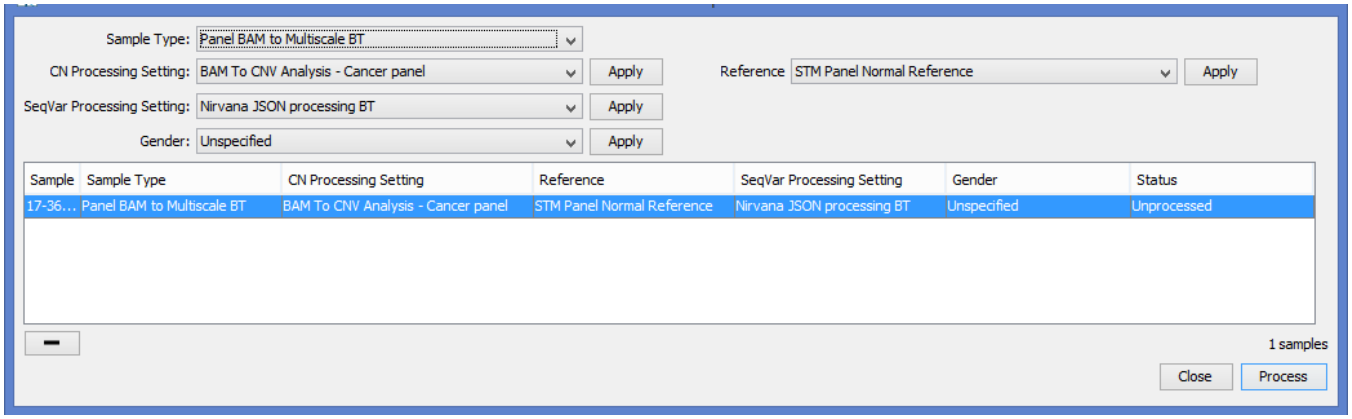


Figure 201. Selection of various settings.

If the sample type has more than one processing type associated with either the CNV or SeqVar modality, the user will need to select which processing settings to use via the dropdown boxes, shown in **Figure 204**. Gender can be left blank. Click **Process**, the status will change to **Pending**, and the window can be closed.

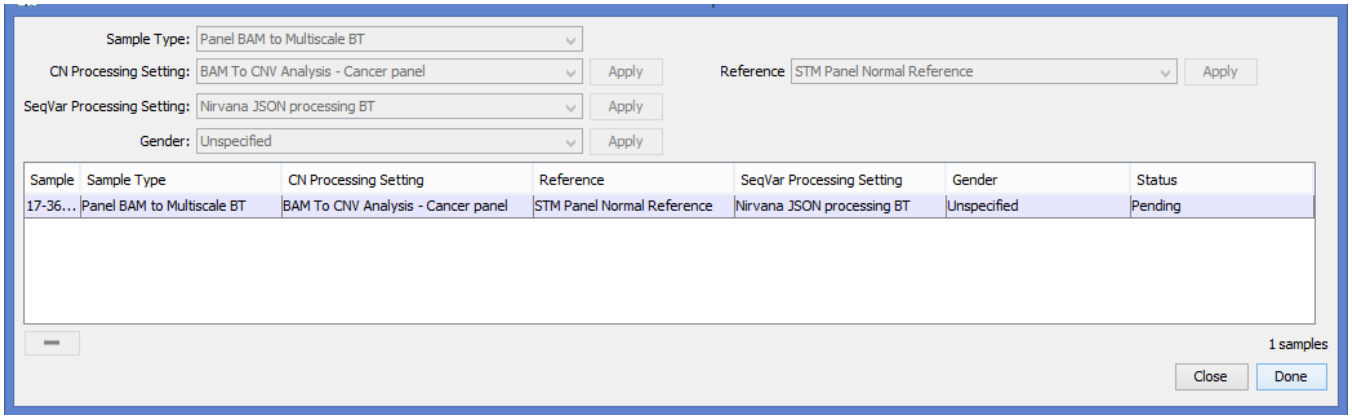


Figure 202. Processing samples.

From the **Home** tab, the sample will appear in the search with status as **Unprocessed** until processing is complete. Once processing is complete the results will be displayed, as shown below in **Figure 205**.

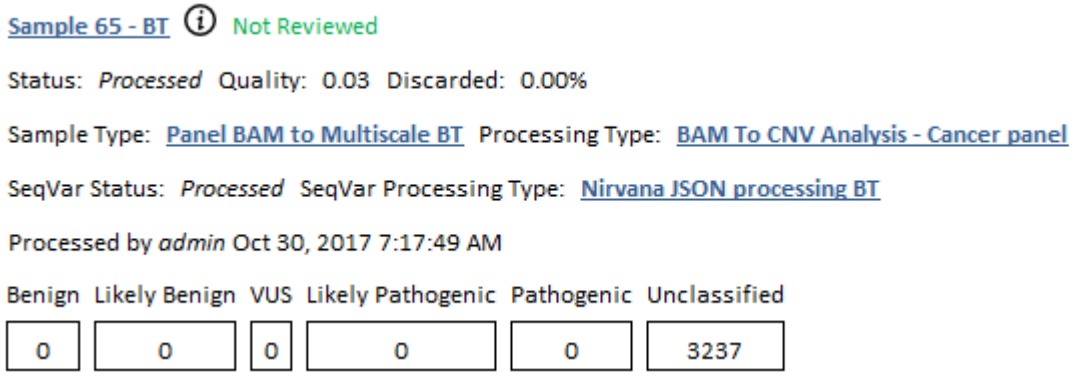


Figure 203. Results from processing.

Knowledge Base

Knowledge Base (KB) stores annotations for any variant or genomic region calls in the software. The KB supports two distinct types of tests – Oncology and Constitutional – containing fields unique to each test type. For Constitutional tests, the KB holds information such as relevant gene overlaps, mode of inheritance, Pubmed IDs and notes for each reference, interpretation, general comments, or classification. For Oncology tests, the KB holds information such as classification based on the AMP-ASCO-CAP guidelines, interpretation, notes, PubMed IDs, and clinical biomarker impact (diagnostic, prognostic, or therapeutic). **NOTE:** The KB currently only supports CNV and AOH events.

The basic process of submission to the KB and approval for both test types is the same, with specific KB content and fields for each test type, as described above. Samples of the Test Type Oncology will only display oncology

KB events and samples of the Test Type Constitutional will only display constitutional KB events. Approved KB events are displayed in the **Variant Details** tab in the **Knowledgebase Events** section.

KB Record for Constitutional Test Type

Upon completing an event submission form, some of the input fields will be automatically pre-populated based on the event's coordinates, type, and overlaps with RefSeq genes. **Figure 206** depicts an example of a KB entry for a CN event for the Constitutional Test Type.

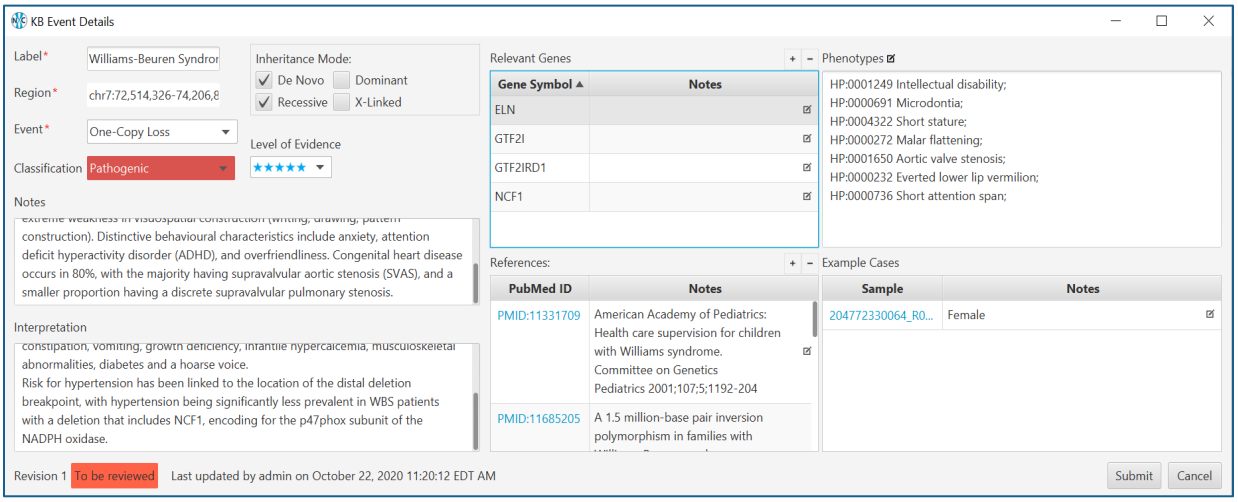


Figure 206. KB record for Constitutional Test Type

The following are windows in the KB Event Details panel that can be edited:

- **Relevant Genes**: gene symbols and notes can be added or subtracted in the pop-up window; highlight one or more genes and click.
- **Level of Evidence**: ranges from one to five stars.
- **Classification**: lists values defined within the event's Sample Type. These values may differ for different Sample Types.
- **Example Cases**: fields will automatically be populated with the submitted event's Sample Name; Notes may be added manually.
- **References**: any bibliographic references may be added using PubMed IDs which will be hyperlinked to PubMed.

KB Record for Oncology Test Type

Figure 207 shows an example of a submission to the KB window for an Oncology KB event and **Figure 208** depicts Oncology KB Event details for an approved event. These windows contain the following features, in addition to some of the ones described above:

- **Clinical Impact**: select a clinical significance from the AMP-ASCO-CAP classifications (J Mol Diagn. 2017 Jan;19(1):4-23).

- **Cancer Types**: select one or multiple entries from WHO and/or OncoTree ontologies. An explicit text search engine is also available to filter down query results.

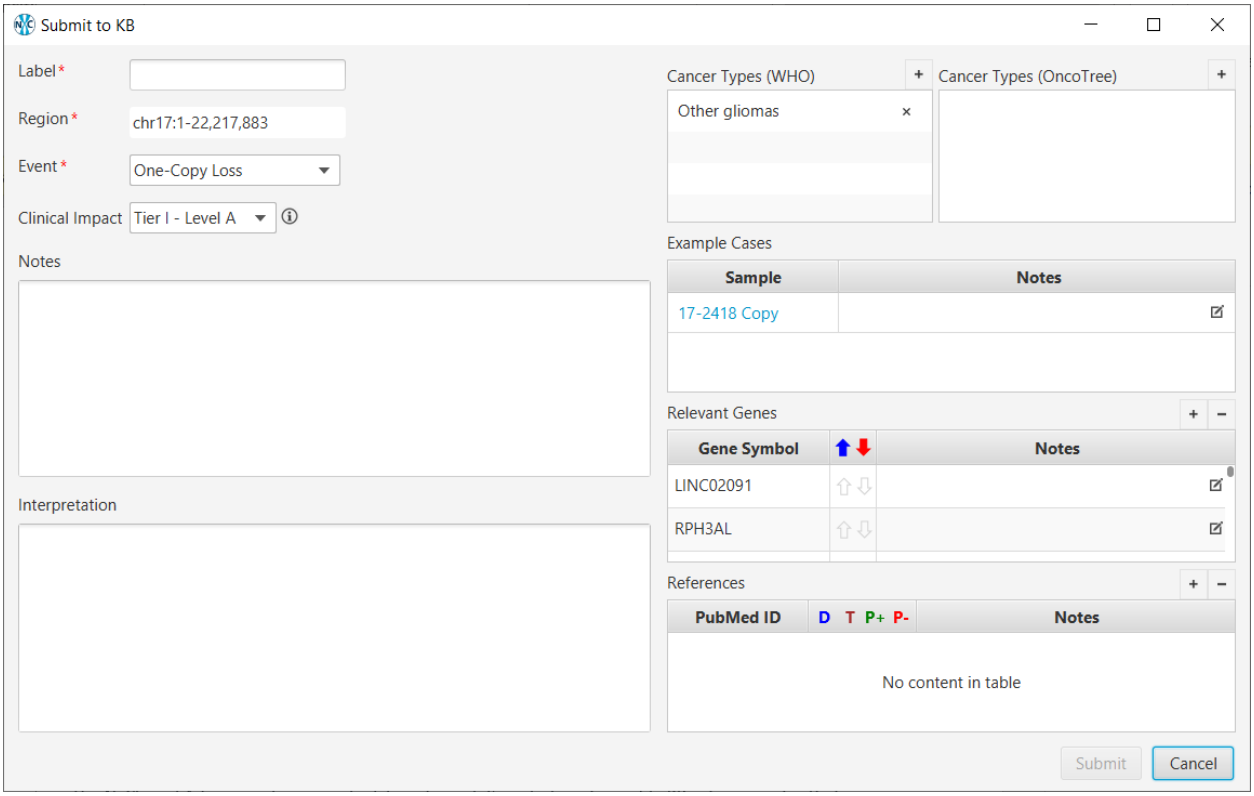


Figure 207. KB submission for Oncology

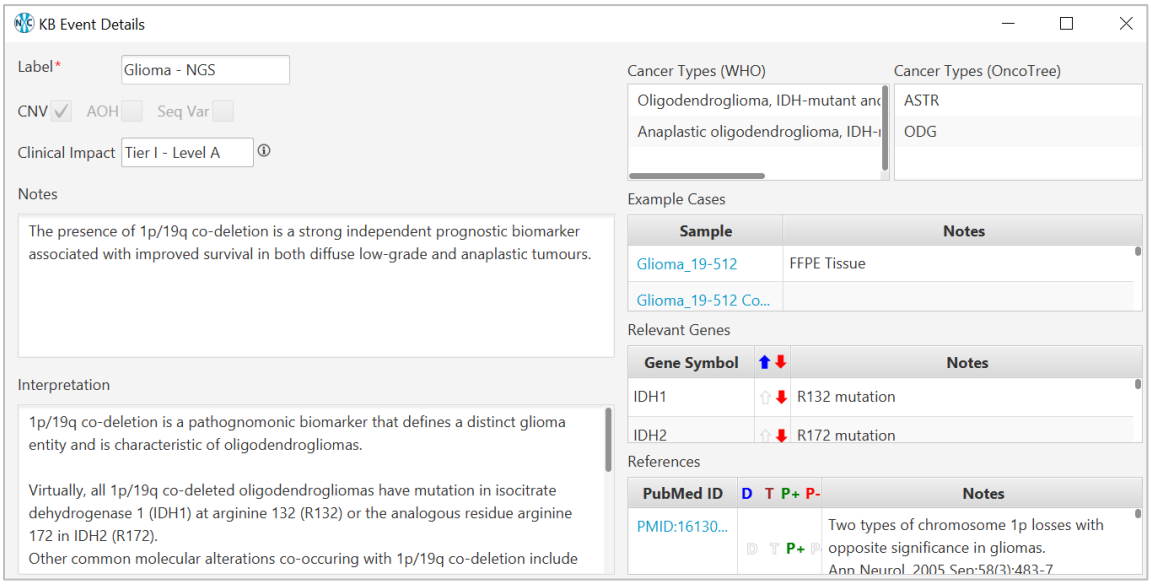


Figure 208. Oncology KB Event details for an approved event

Admin Features Related to the KB

TEST TYPE

The Admin must assign a test type to each sample type to enable KB submission for that sample type, shown in **Figure 209**. By default, this field is blank. Sample types created in version 5.2 and older will have no value for this field. To enable full KB features on these older samples, the Admin must assign a test type for each sample type for which KB submission should be enabled.

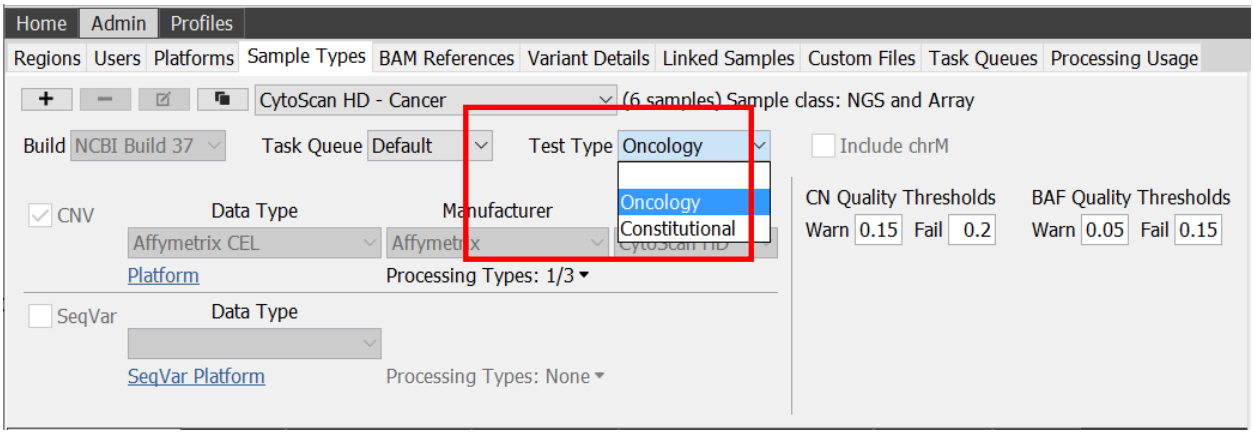
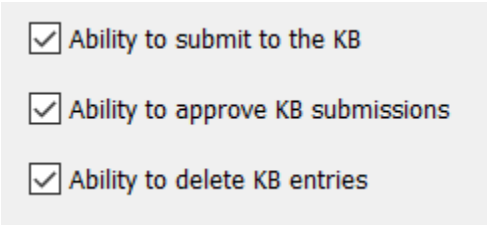


Figure 209. Assigning a Test Type by the Administrator

For backwards compatibility, VIA defaults any sample type with a blank test type to a **Constitutional** workflow, enabling a user to display **Constitutional KB** events in both the **KB** tracks and the **Variant Details** tab. VIA also enables any user with relevant KB permissions to edit, approve and reject KB submissions for the **Constitutional** workflow only. Submitting events to the KB is disabled for sample types without a defined test type.

USER PERMISSIONS

Available permissions granted by the Admin are:



Oncology Profiles

Profiles and Aggregates

Users can leverage cancer samples and information stored in the unique knowledge base by pooling these samples together to create CNV profiles. These profiles/signatures can then be ranked on similarity to new cases to assist with interpretation.

These cancer signatures can be created based on the samples within VIA, but a signature can also be loaded from an external source.

CNV profiles can be used as prognostic markers, to identify the origin of classification of the tumors, or even to assist with diagnosis.

Profiles are created first by generating an aggregate displaying the CNV/AOH frequency of a set of samples in the database selected based on cancer classification (WHO/OncoTree) and any attributes (e.g., cancer subtype, tissue).

The aggregate is then converted into a profile by adding annotations (clinical impact, list of actionable genes, PubMed IDs) and saved in the KB. Stored cancer profiles in VIA are akin to events added to the KB; the main difference is that they are representative of several samples rather than a single sample, as cancer is typically a multi-gene and genome-wide phenomenon.

Once a profile is submitted for inclusion in the KB, it follows the approval protocol. Approved profiles in the KB can be updated in the future with new curated samples by accessing the associated aggregate.

Aggregate versus Profile

An aggregate is a dynamic representation of a set of samples matching certain criteria as expressed by a query. The aggregate changes and reflects the status of the query based on the current set of samples. As sample sets grow or existing samples obtain specified or edited attributes additional samples may meet the query criteria or now fail the criteria changing the aggregate.

A profile is a snapshot of the aggregate at a specific point in time and is also annotated.

NOTE: The KB currently only supports **CNV**, **AOH** and **SeqVar** events.

Aggregates and Profiles

An aggregate displays the frequency of **CNV** or **Allelic** events across a set of samples and is an intermediate in the process. It may be generated by using samples within VIA or by importing an external file. An aggregate is a snapshot of the frequency profile of a set of selected samples at a specific time. The aggregate does not get updated automatically upon addition or deletion of samples within VIA, but it may be updated to reflect new samples in the database or ones that have been deleted by editing and re-saving the aggregate.

Creating an aggregate: In the **Profiles** tab, click the **+** icon, as shown in **Figure 210**. The **Search** function can be used for sample names, and wildcards are accepted (e.g., **_cancer* will return all samples that end in _cancer). Only samples with **Test Type=Oncology** will be included in profiles. Additional filtering criteria may be specified using the filter categories available, such as sample type. **Aggregate Name** is required but **Description** is optional. See **Figure 211** for a display of these features.

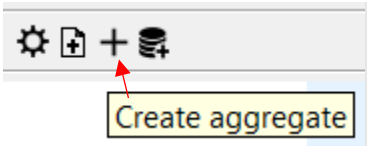


Figure 20410. Creating an Aggregate.

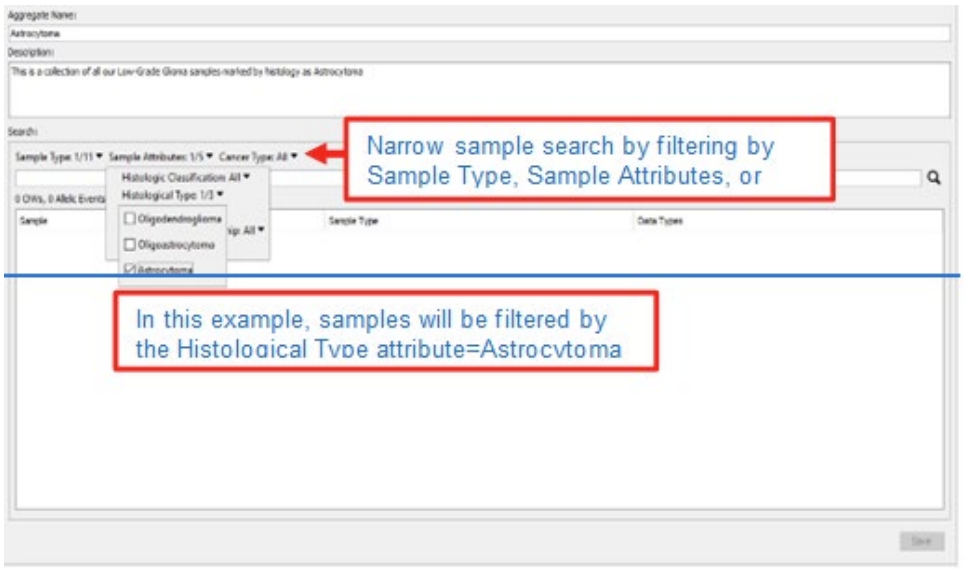
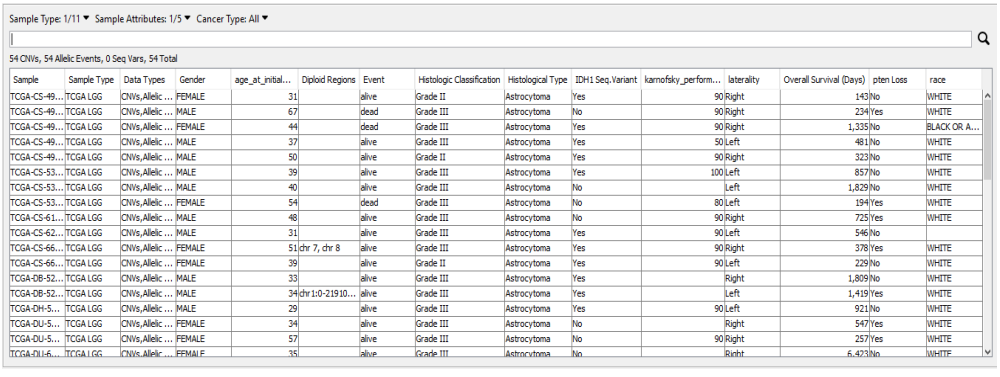


Figure 20511. Aggregate features.

Matching samples (processed) will be listed along with all attributes after executing the search, as shown in **Figure 212**.



Sample	Sample Type	Data Types	Gender	age_at_initial...	Diploid Regions	Event	Histologic Classification	Histological Type	IDH1 Seq.Variant	karnofsky_perform...	laterality	Overall Survival (Days)	pten Loss	race
TCGA-CS-49...	TCGA LGG	CNVs,Allelic ...	FEMALE	31		alive	Grade II	Astrocytoma	Yes	90	Right	143	No	WHITE
TCGA-CS-49...	TCGA LGG	CNVs,Allelic ...	MALE	67		dead	Grade III	Astrocytoma	No	90	Right	234	Yes	WHITE
TCGA-CS-49...	TCGA LGG	CNVs,Allelic ...	FEMALE	44		dead	Grade III	Astrocytoma	Yes	90	Right	1,335	No	BLACK OR A...
TCGA-CS-49...	TCGA LGG	CNVs,Allelic ...	MALE	37		alive	Grade III	Astrocytoma	Yes	50	Left	481	No	WHITE
TCGA-CS-49...	TCGA LGG	CNVs,Allelic ...	MALE	50		alive	Grade II	Astrocytoma	Yes	90	Right	323	No	WHITE
TCGA-CS-53...	TCGA LGG	CNVs,Allelic ...	MALE	39		alive	Grade III	Astrocytoma	Yes	100	Left	857	No	WHITE
TCGA-CS-53...	TCGA LGG	CNVs,Allelic ...	MALE	40		alive	Grade III	Astrocytoma	No		Left	1,829	No	WHITE
TCGA-CS-53...	TCGA LGG	CNVs,Allelic ...	FEMALE	54		dead	Grade III	Astrocytoma	No	80	Left	194	Yes	WHITE
TCGA-CS-61...	TCGA LGG	CNVs,Allelic ...	MALE	48		alive	Grade III	Astrocytoma	No	90	Right	725	Yes	WHITE
TCGA-CS-62...	TCGA LGG	CNVs,Allelic ...	MALE	31		alive	Grade III	Astrocytoma	Yes	90	Left	546	No	
TCGA-CS-66...	TCGA LGG	CNVs,Allelic ...	FEMALE	51	chr 7, chr 8	alive	Grade III	Astrocytoma	Yes	90	Right	378	Yes	WHITE
TCGA-CS-66...	TCGA LGG	CNVs,Allelic ...	FEMALE	39		alive	Grade II	Astrocytoma	Yes	90	Left	229	No	WHITE
TCGA-DB-52...	TCGA LGG	CNVs,Allelic ...	MALE	33		alive	Grade III	Astrocytoma	Yes		Right	1,809	No	WHITE
TCGA-DB-52...	TCGA LGG	CNVs,Allelic ...	MALE	34	chr 1:0-21910...	alive	Grade III	Astrocytoma	Yes		Left	1,419	Yes	WHITE
TCGA-DH-5...	TCGA LGG	CNVs,Allelic ...	MALE	29		alive	Grade III	Astrocytoma	Yes	90	Left	921	No	WHITE
TCGA-DU-5...	TCGA LGG	CNVs,Allelic ...	FEMALE	34		alive	Grade III	Astrocytoma	No		Right	547	Yes	WHITE
TCGA-DU-5...	TCGA LGG	CNVs,Allelic ...	FEMALE	57		alive	Grade III	Astrocytoma	No	90	Right	257	Yes	WHITE
TCGA-DU-6...	TCGA LGG	CNVs,Allelic ...	FEMALE	35		alive	Grade III	Astrocytoma	No		Right	6,423	No	WHITE

Figure 20612. Matching sample list.

Click **Save** to create the aggregate. The query (with filters applied) is saved, not actual samples. When new samples are processed and they match the query, the aggregate is automatically updated to include the new

samples. If this addition of samples occurs while an aggregate is currently open and displayed in the track, it must be closed and re-opened to reflect the new samples.

Displaying Aggregates and Profiles

Display, deletion, and editing of Aggregates and Profiles is managed in the same UI. Once created, edit the visibility of the Aggregate or Profile to view it in the **Profiles** window. Click on the **Gear** icon to open the **Aggregate/Profiles Selection** window, demonstrated in **Figure 213**.

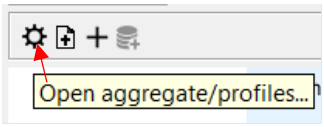


Figure 20713. The **Gear** icon opens the selection window.

Mark checkboxes to display profiles/aggregates in the **Profiles** tab. The **Track Layout** section, shown in **Figure 214** at the bottom of the window, is used to order the plots. Click and drag names up and down to arrange them into a desired layout. **Aggregates** and **Profiles** are in the **Track Layout** section:

- The **Track** modality is displayed in parentheses next to the track name (e.g., CNV Events).
- If a Profile's associated Aggregate has the same name, they can be distinguished from each other.
  - Profiles have labels: **Pending** or **Approved**.
  - Aggregate names do not have labels.

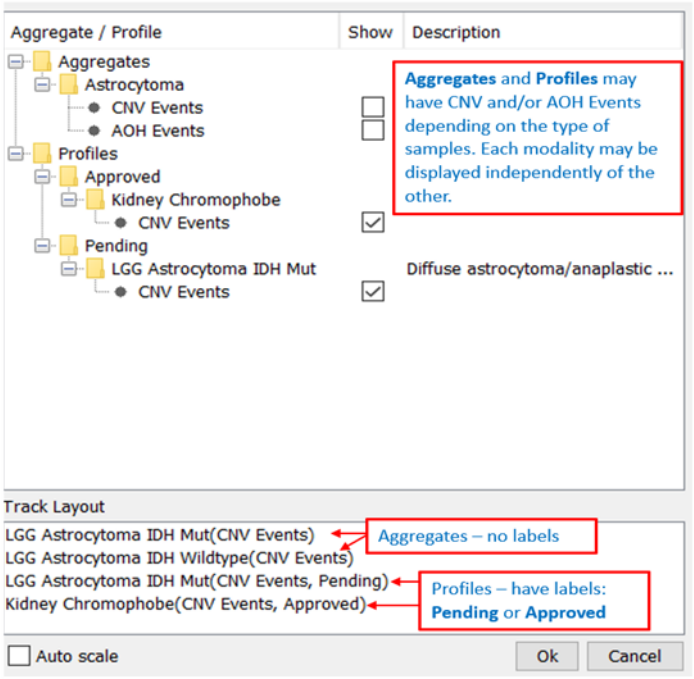


Figure 2084. Track layout.

Auto scale: When not enabled, plots are displayed on the Y-axis scale from 0 to 100%. With the checkbox marked, the Y axis automatically scales to the Y-axis max (each track track's highest frequency). See **Figure 215** and **Figure 216**. Close or hide a track by clicking the **Close** (**x**) button in the top left of the track.

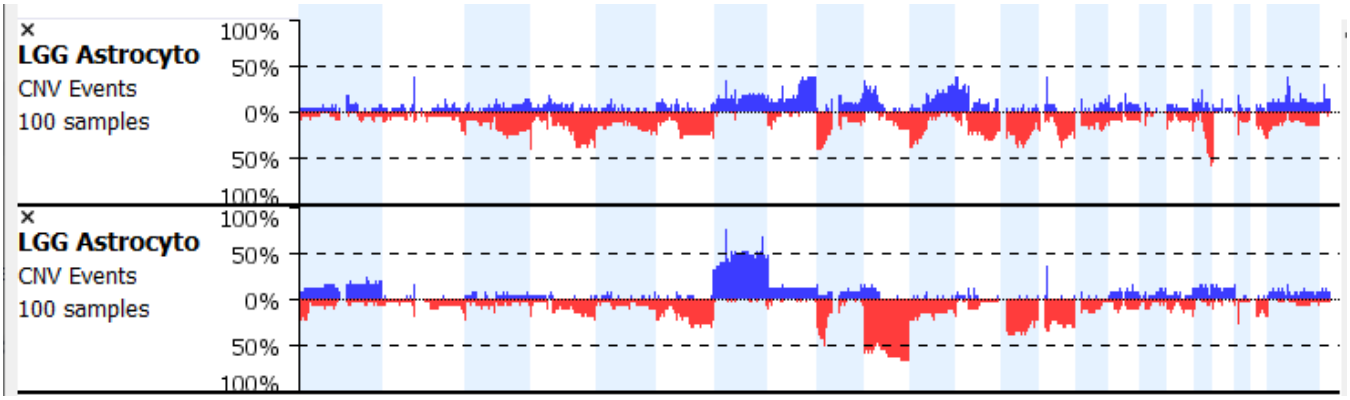


Figure 2095. Auto scale not enabled (box unchecked).

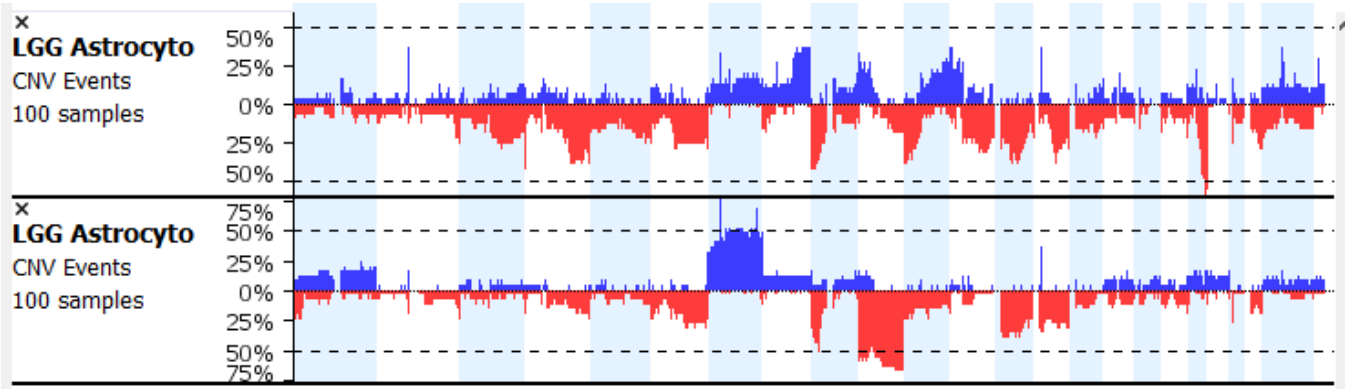


Figure 2106. Auto scale enabled (box checked).

Editing Aggregates and Profiles

Different tools are available for each track based on type, and can be used for editing purposes, as shown in **Table 15** and **Table 16**.

Table 15. Different editing tools.

Remove Remove the track from VIA	**Edit** Make changes to this track	**Refresh** Refresh/update with new aggregate data	**Approve** Approve a profile in **Pending** state

Table 16. Tools available based on track type.

TRACK TYPE	AVAILABLE TOOLS
Aggregate loaded as bedGraph file	None as it is temporary and only available for the current user session in which it is loaded.
Aggregate created using samples in the db	Edit Remove

Profiles	Edit Refresh Remove Approve (Profile with Pending status only)

Refresh opens the **Select Profile Data** window which lists the currently displayed aggregates and one from the list can now be associated with the profile. A profile can be updated using the original aggregate employed to generate the profile or a different aggregate can be associated with the profile. **Figure 217** and **Figure 218** are shown as examples of these editing capabilities.

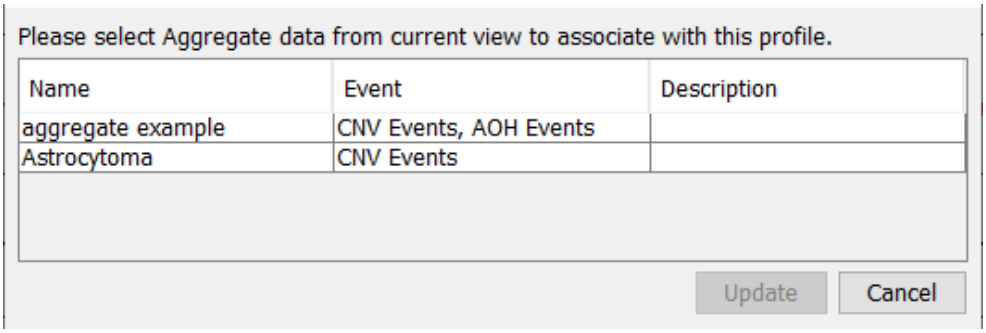


Figure 2117. Updating a Profile.

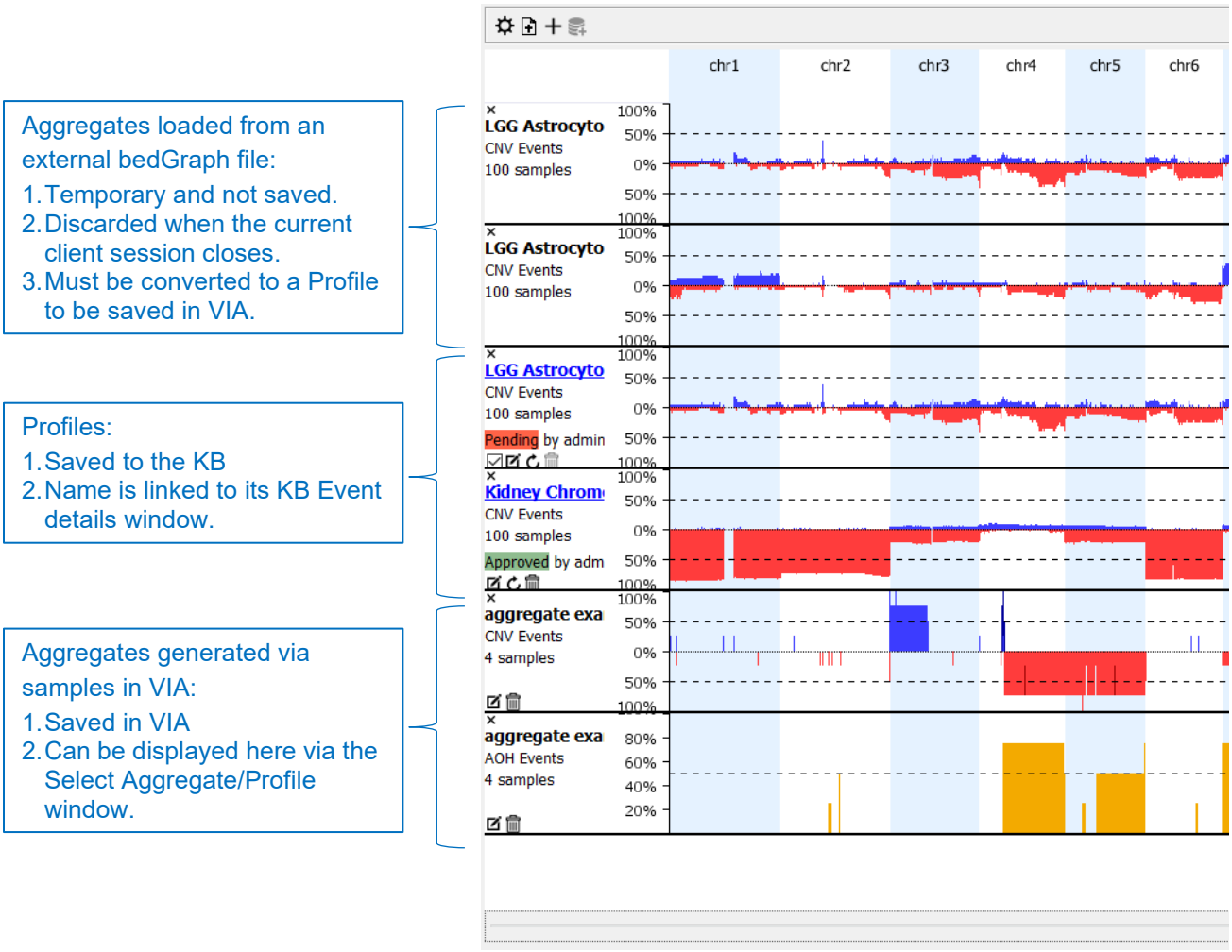


Figure 2128. An updated Profile.

When editing an aggregate, the window opens listing all current samples matching selected filters and queries.

- Listed samples may be greater or fewer than the number of samples in the current Aggregate as the Aggregate is dynamic and changes as sample lists grow.
- The query may be modified and must be saved again to update the aggregate. The criteria may be altered to include samples of another sample type or those matching additional attributes. After altering the selection criteria, click the **Search** button to list sample results. Then click **Save** to take a new snapshot and update the Aggregate.

Converting an Aggregate to a Profile and Submitting to the KB

An oncology profile is a snapshot of the Aggregate including annotations (see **Figure 219**) and is stored in the KB. To add an aggregate to the KB as a profile, click in the area with the aggregate name to highlight in green.

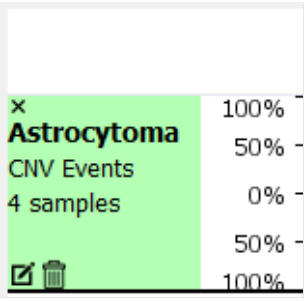


Figure 2139. Oncology snapshot.

The **Submit to KB** button, shown in **Figure 220**, becomes active (no longer gray). Click on the button to bring up the **Submit to KB** window and annotate the profile by filling out the fields in the **Profile** window. See the **Knowledge Base** section for details on how to fill out the **KB** record. Follow the KB submission and approval steps to add the **Profile** to the **KB**.

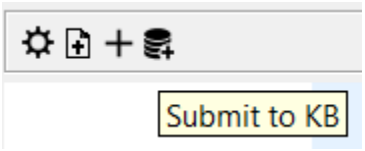


Figure 21420. KB button is no longer gray.

Loading an Aggregate from a File

Aggregates can also be loaded from an external bedGraph (.bgr) file, shown in **Figure 221**. The file contains tracks with frequency of gains and losses at specified chromosomal locations. After loading the **BGR** file tool and selecting the BGR file, an input window opens, as shown in **Figure 222.**

track type=bedGraph name="LGG Astrocytoma IDH Wildtype gains"					
chr1	62536	585194	4		
chr1	585194	751456	8		
chr1	751456	3372743	12		
track type=bedGraph name="LGG Astrocytoma IDH Wildtype losses"					
chr10	46207226	46981565	-48		
chr10	46981565	47055683	-52		

Figure 21521. Loading from external files.

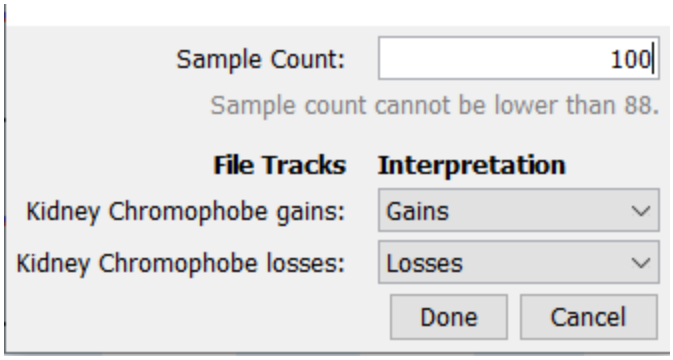


Figure 21622. Input window for BGR files.

For BGR files, keep the **Sample Count** field at 100 (indicating percentage of samples, not the actual count of samples), shown in **Figure 223**. As the values in the files are the frequency of gains/losses, the values have a max of 100. Once loaded, select the track to be displayed before it shows up in the view.

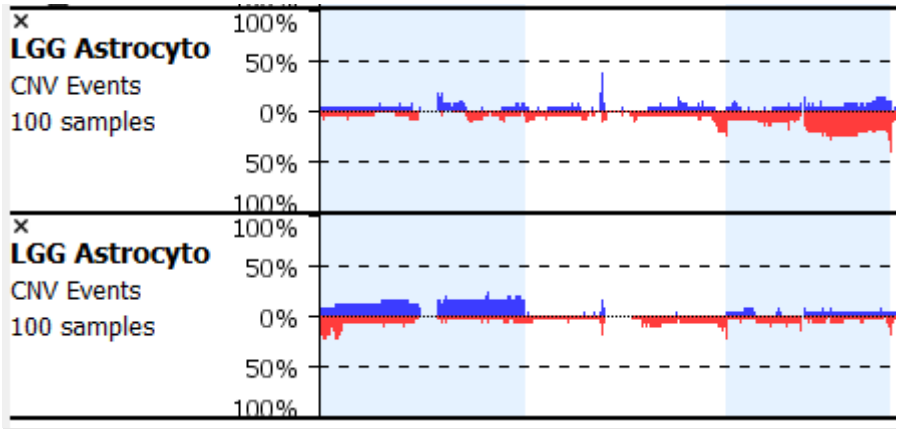


Figure 21723. Sample count at 100.

Aggregates loaded from an external bedGraph file will not have tools to edit or delete as there are no corresponding samples in VIA for that aggregate.

The Aggregate may be submitted to the KB and knowledge from published papers can be incorporated with a Profile. For example, the section of **Table 17** from the Compendium of Cancer Genome Aberrations has information about Diffuse Astrocytoma. The information here can be used to annotate an externally loaded LGG Astrocytoma profile into the KB, as seen in **Figure 224**.

Table 17. Compendium of Cancer Genome Aberrations.

Infiltrating Gliomas	Diffuse astrocytoma/anaplastic Astrocytoma, WHO grade II/III, IDH mutant	**Gain:** 4q, 7q, 8q24, 12q **Loss:** 9p, 19q (without 1p)	**Gain:** MYC **Loss:** CDKN2A/B, PML15q22	Better prognosis than IDH wildtype astrocytoma; Progression to grade IV will often involves loss of 10q, gain of CDK4, CDK6, and cyclin E2, and an increase in copy number alterations.	PMID:26824661; PMID:26061753; PMID:25263767 PMID:26061754; PMID:28535583; PMID:26091668 PMID: 25701198; PMID:26865861; PMID:29687258
	Diffuse astrocytoma/anaplastic astrocytoma, WHO grade II/III, IDH wild-type	**Gain:** 7, 19 **Loss:** 4, 9p 10 **Amplification:** EGFR, MDM4, CDK4	**Loss:** homozygous CDKN2A/B **Mutation:** EGFR, NF1, PTEN **Amplification:** EGFR, MDM4, CDK4	Poor prognosis with similar abnormalities to glioblastoma	PMID:26061754; PMID:26824661;PMID:28535583 PMID:26091668; PMID:26810070

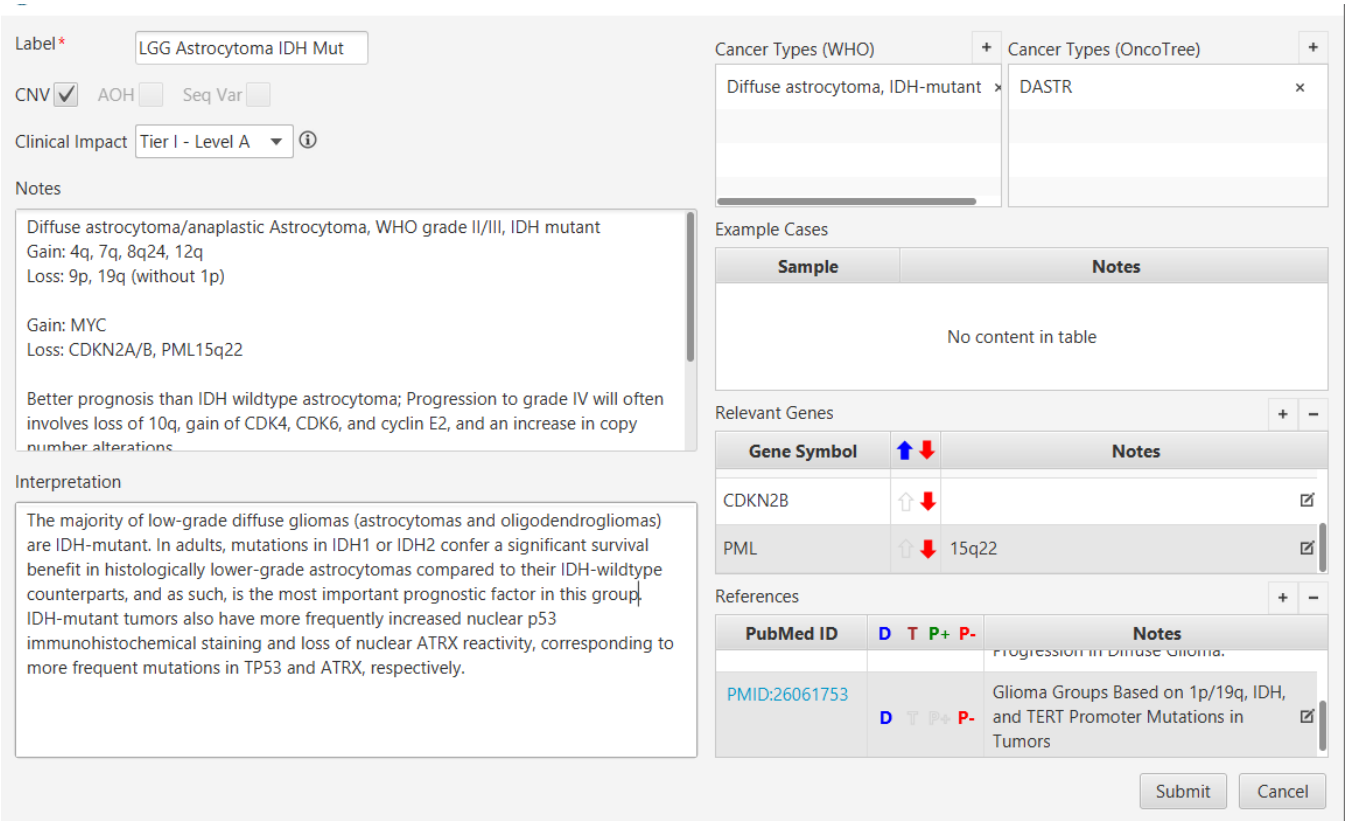


Figure 2184. The LGG Astrocytoma IDH Mutant profile is annotated with information from **Table 17** above.

Detecting CNV from NGS

Detecting CNV and AOH events from NGS data using the MultiScale Reference (MSR) and Self-Reference algorithms affords flexibility when calling from different NGS assays such as Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS), as shown in **Figure 225**.

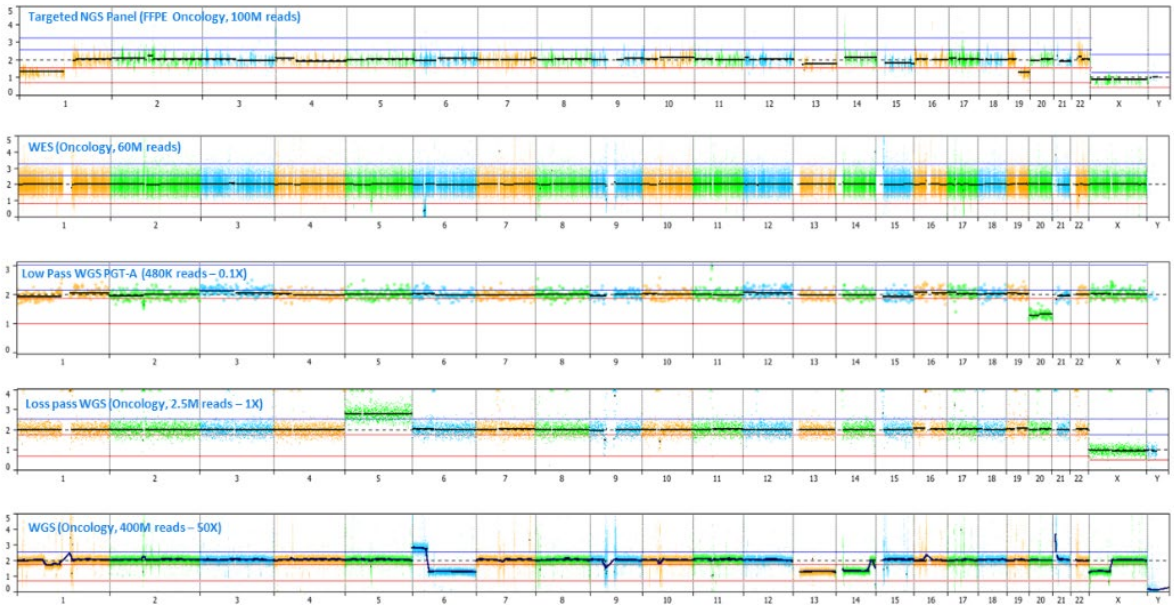


Figure 2195. Example of Copy Number Analysis – Portfolio.

BAM MSR requires a reference file created from a pool of normal diploid samples and is compatible with NGS enrichment assays (WES, panels). Performing various systematic corrections (e.g., GC bias), dynamic binning also obtains higher resolution on capture areas and lower resolution for backbone (off-target) reads.

- Recommended: 10-15 normal samples of same sex.
- Samples can be normal or from an experimental set which does not share CNV events.
- Users may need to create multiple reference files using diff. input files or change parameters to find one that works best.
- Users will need to create and load reference before samples can be processed and the **MultiScale Reference Builder** (see **Figure 226**) needs to be installed separately for ease of use.

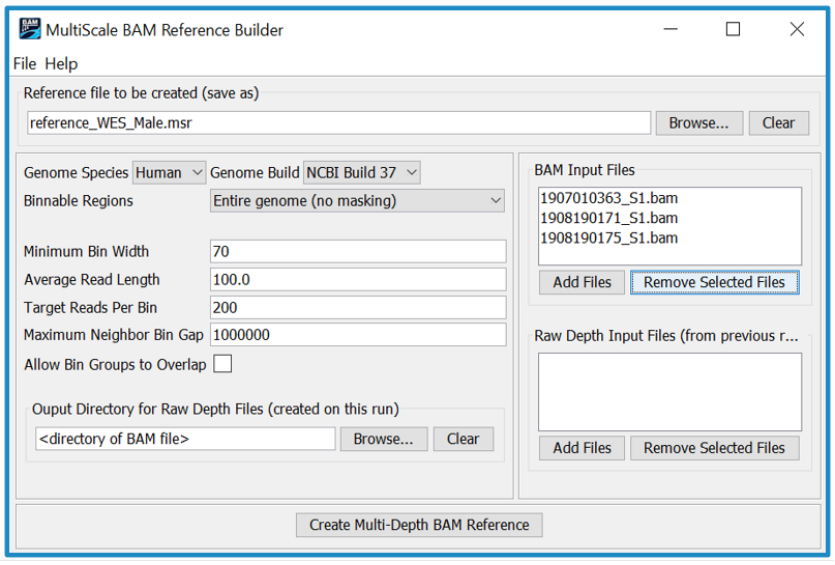


Figure 2206. Example of typical settings for WES.

Parameters for BAM MSR

- **Minimum Bin Width**: The smallest contiguous region to bin; smallest possible resolution.
- **Average Read Length**: Based on technology/platform used.
- **Target Reads Per Bin**: Reads to tally in a bin.
- **Maximum Neighbor Bin Gap**: Maximum number of nucleotides allowed between neighboring bins; can be used to prevent spanning the undefined sequence at the centromere and merge data far apart in the genome.
- **Binnable Regions (optional):** Generally, for WES and large gene panels, it is preferable to also use the off target reads to create virtual probes; this will generate a lower resolution backbone. Select: **Entire genome (no masking)** for this option. In the case of small NGS targeted panels (ten genes), it is preferable to customize binnable regions. Select: **Custom** (use a .BED or .TXT masking file with the first three columns as chromosome, start, stop) for small panels.
- **Figure 227** below represents the parameters described above.

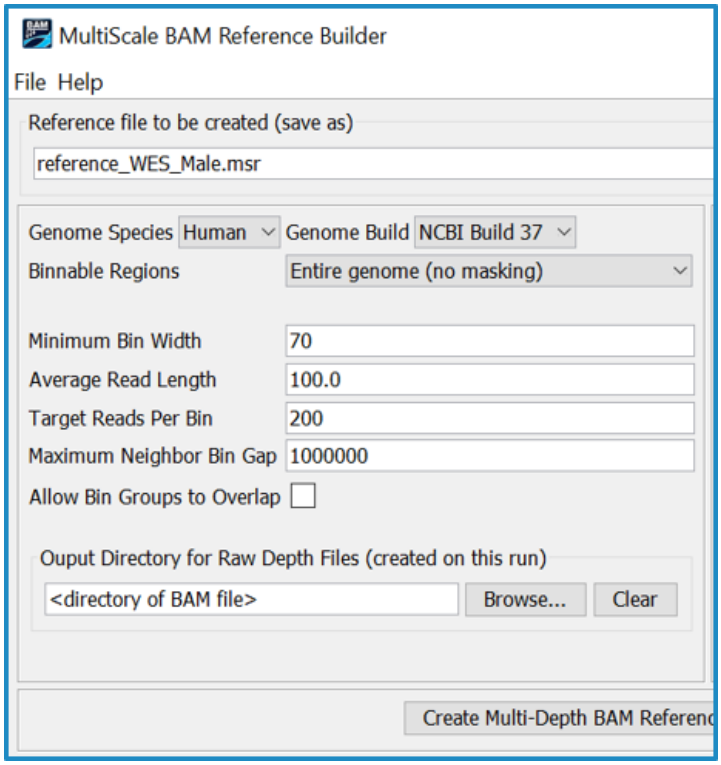


Figure 2217. Parameter indications for the MSR BAM Reference Builder.

Additional parameter information includes the following:

Raw Depth Input files (from previous runs) are intermediate binary files created for each BAM file after it is processed, which can be ignored if creating a reference file for the first time. In other words, use when the reference file needs to be re-built or when adding additional files to an existing reference. **Figure 228** illustrates this utility.

Settings for **Genome Species**, **Genome Build**, **Custom Binnable Region File** and **Minimum Bin Width** must be the same as those used to generate the depth files.

Considering the **Output** directory for Raw Depth files (created on this run), the location must be specified to store depth files in a separate directory from where the BAM files are located; otherwise, depth files will be created in the same directory as the respective BAM file.

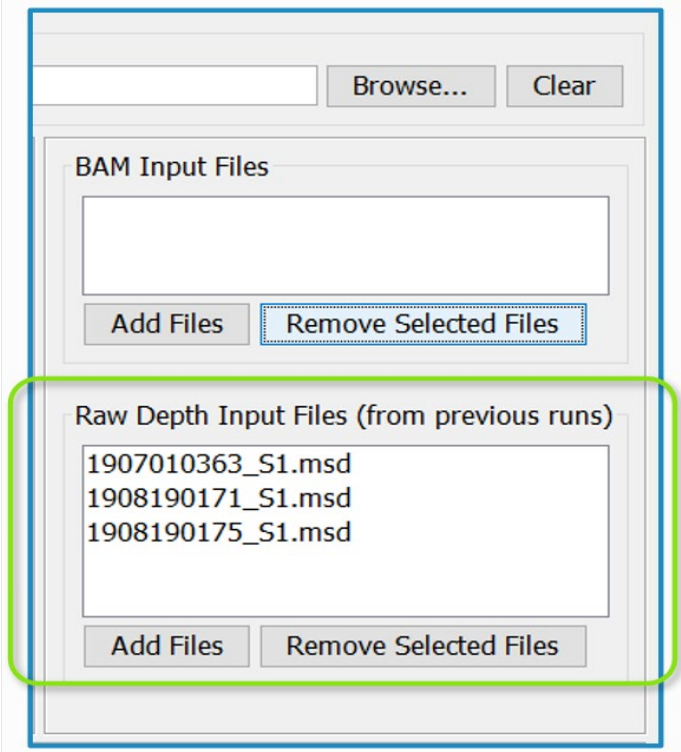


Figure 2228. Raw Depth Input files from previous runs.

Allow Bin Groups to Overlap – Check the box to allow surrounding low density bin groups to merge across (overlap) the high-density singleton bin groups. Minimum width bins that have at least the target nucleotide count are converted directly to bin groups comprising a single minimum width bin. To create a BAM MSR reference file, see **Figure 229**.

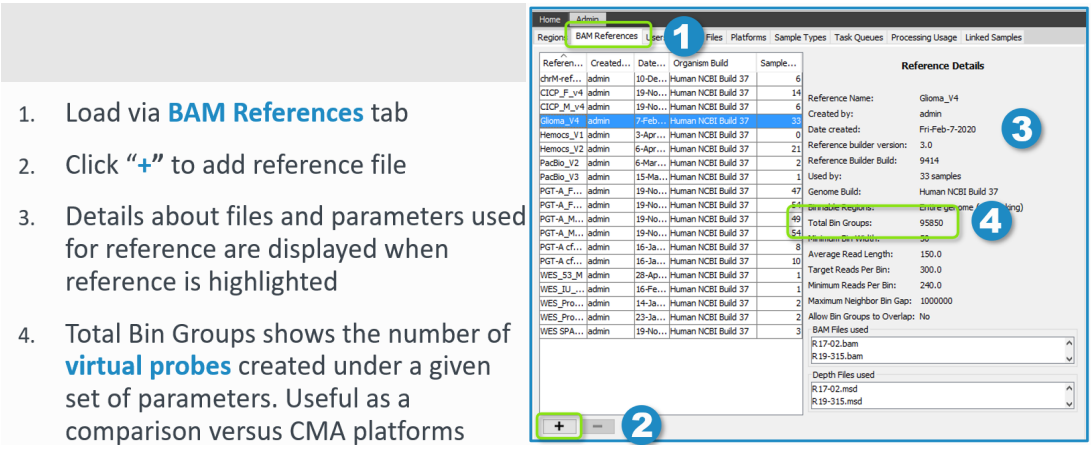


Figure 2239. Steps for creating a reference file that includes graphical representation.

Generating BAF values from BAM files

- Details under **Platforms > BAM Multiscale > Processing Types** (see **Figure 230**).
- Creates a BAF value from the BAM pile-up for known SNP positions (SNP file for BAF).
- Each candidate's SNP position is checked for minimum read depth, alignment, and base quality.
- Quality and density of BAF measurement is assay specific.

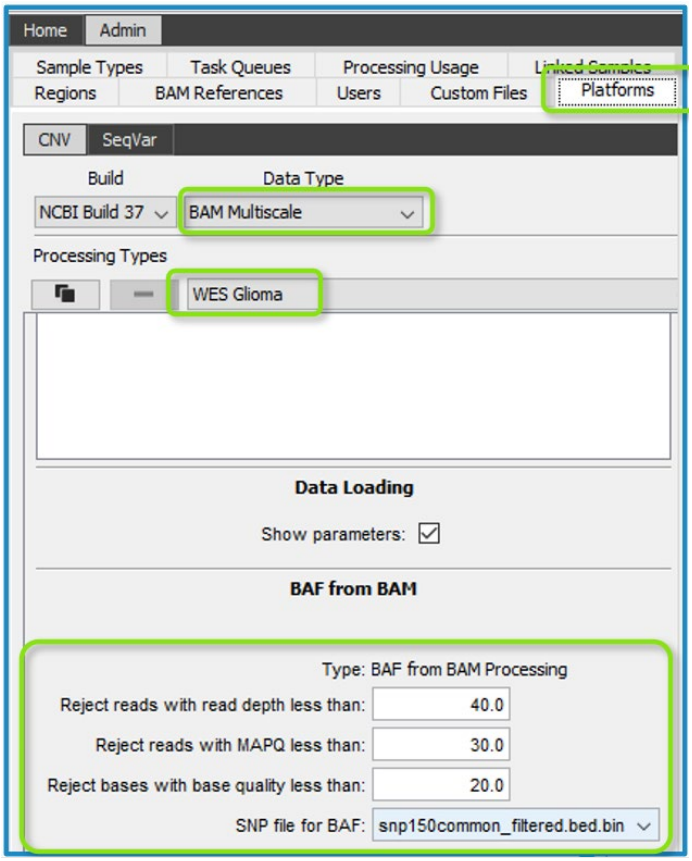


Figure 22430. Generating BAF values from BAM files.

Copy Number Analysis

The MSR example in **Figure 231** shows the relationship of bins density to capture regions in WES samples, illustrating how bins are dynamically generated depending on depth of coverage.

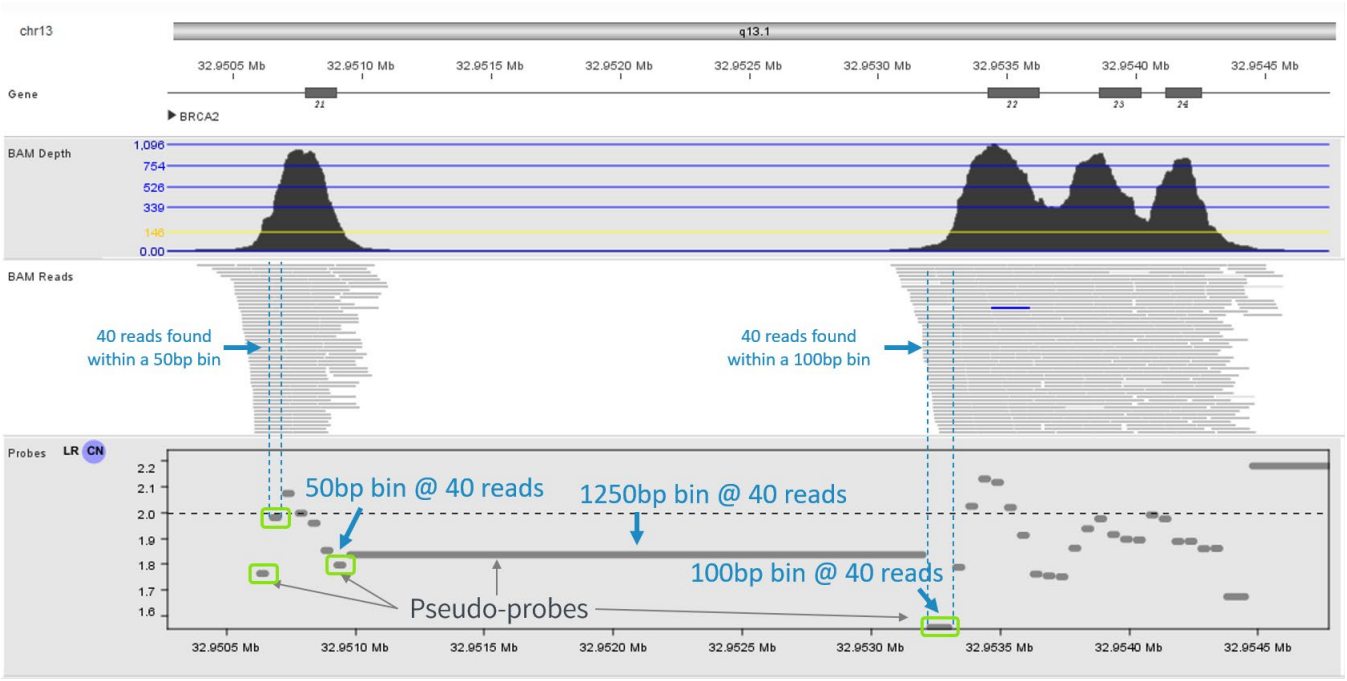


Figure 225. Copy Number Analysis (BAM-MSR).

Figure 232 is an example of a heterozygous deletion generated from the BAM MSR algorithm. **TOP:** Genome-wide copy-number and B-Allele Frequency plots displaying the generated bins from the BAM MSR algorithm. **BOTTOM:** Detailed view of a 17q11.21 microdeletion showing both a copy-number plot from the BAM MSR - generated bins (Probes) and its allelic confirmation from the B-Allele Frequency plot.

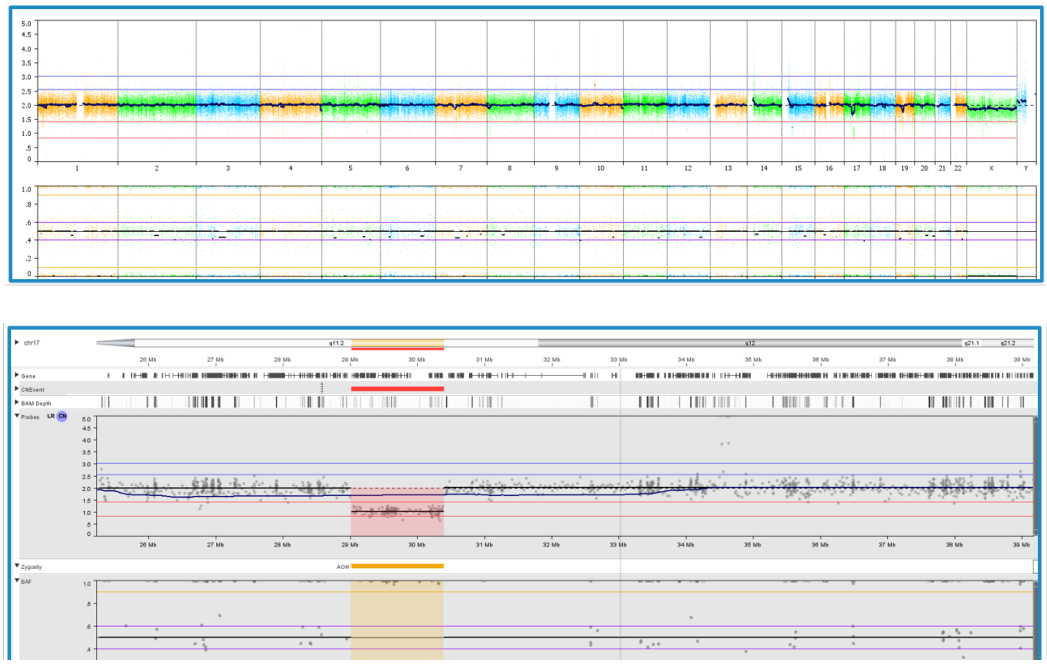


Figure 22632. Graphical views of BAM MSR data.

BAM Self-Reference works on single BAM files generated by WGS, dividing the genome into regular bins to count reads. Self-reference calculates the median read depth per bin and uses it to normalize all bins, performing various systematic corrections (e.g., GC Bias) and masking out certain variable regions in the genome, in the process.

- Designed for WGS assays.
- Settings can be found under **Platforms ➔ BAM Self Reference ➔ Processing Type**.
- Provides recommended bin width based on read depth, as seen in **Figure 233.**
- **Figure 234** describes masking options for variable regions.
- **Figure 235** and **Figure 236** are examples of Self-Reference analysis and coverage.

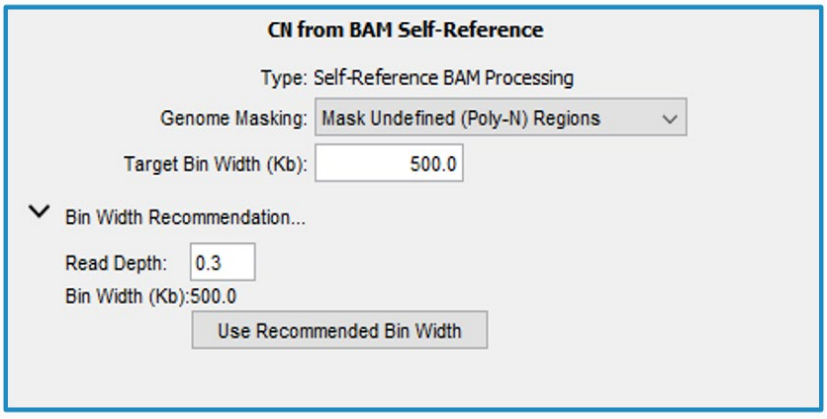


Figure 22733. Bin width window.

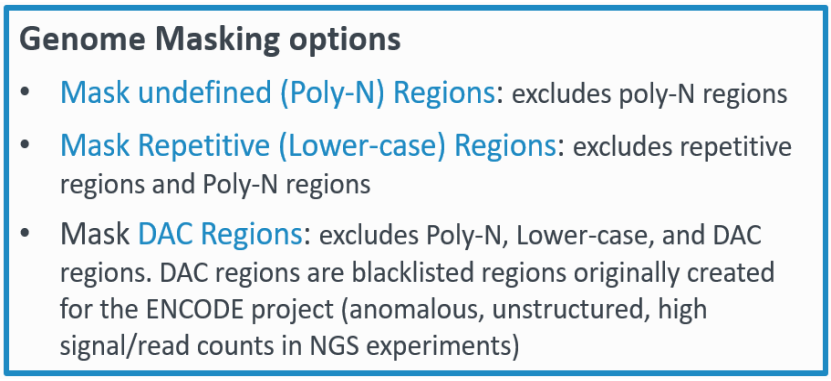


Figure 2284. Different masking options for different regions.

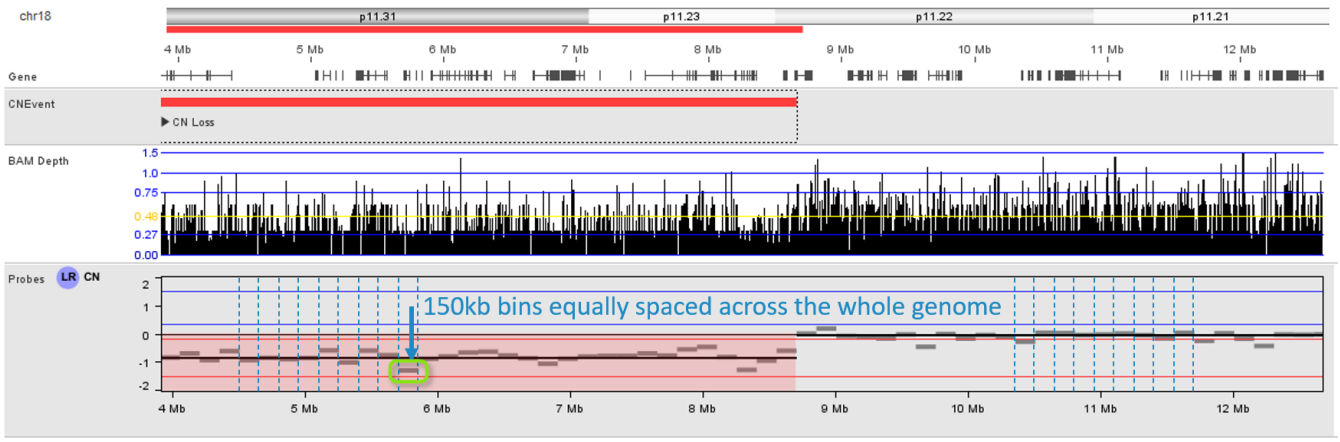


Figure 2295. Copy number analysis: Loss-pass WGS (1X) on a CLL sample.

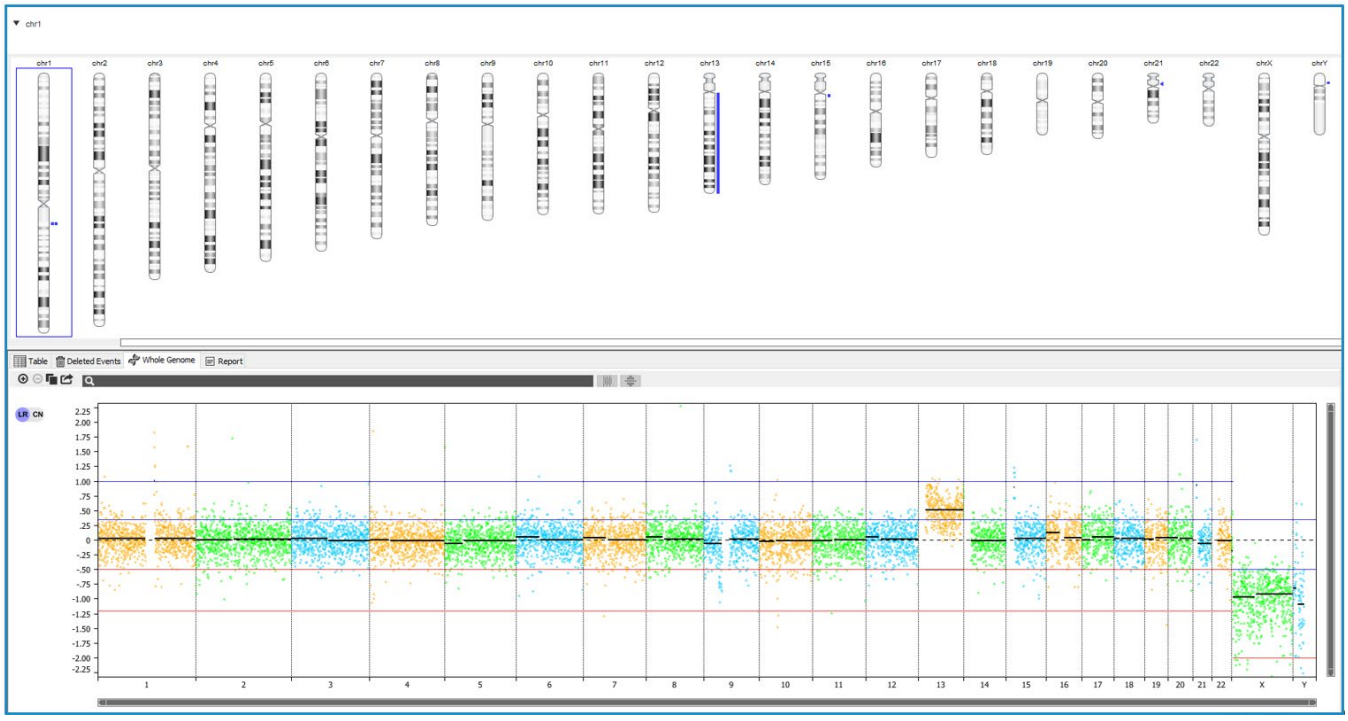


Figure 2306. 250kb bin, 0.7X coverage for Loss-pass WGS (0.1X) on a cell-free DNA assay.

Detecting CNV from Illumina EPIC Methylation Arrays

Illumina Methylation arrays are designed to detect levels of methylation at certain CpG islands using two probes. These arrays measure these levels through the intensity of signals, one for the methylated state and another for the unmethylated. The sum of intensity of the two signals can be used to measure the copy number state of a locus by comparing the total intensity of the probe in the test sample against a reference set.

GENERATING A FINAL REPORT FILE FOR REFERENCE SAMPLES

The **Methylation CN Reference Builder** application available from Bionano should be installed to generate the methylation reference files and the reference sample data should be in the Final Report (TXT) format, obtained from Illumina's **GenomeStudio** 2011. The Final Report file must minimally include data fields for TargetID, Signal_A, Signal_B, and Intensity. Users can create a single Final Report file or individual files for each of the reference samples. Refer to Illumina's **GenomeStudio** manufacturer's manual for further guidance on the use of the software to create a grouped **Methylation Project** and to export the **Sample Methylation Profile Final Report** files.

BUILDING A METHYLATION CN REFERENCE FILE

The Methylation CN Reference Builder is a separate utility that can be installed by the user, as shown in **Figure 237**. Within the **Methylation CN Reference Builder** program, browse to the desired location to save the reference file. Next, choose the correct genome build from the dropdown menu choices. If a genome build is not available, contact Bionano Support to obtain files for the preferred genome build. Finally, select **Add Files** to navigate to the **Final Report** files for the reference dataset, then select **Create Methylation Reference** to initiate the program.

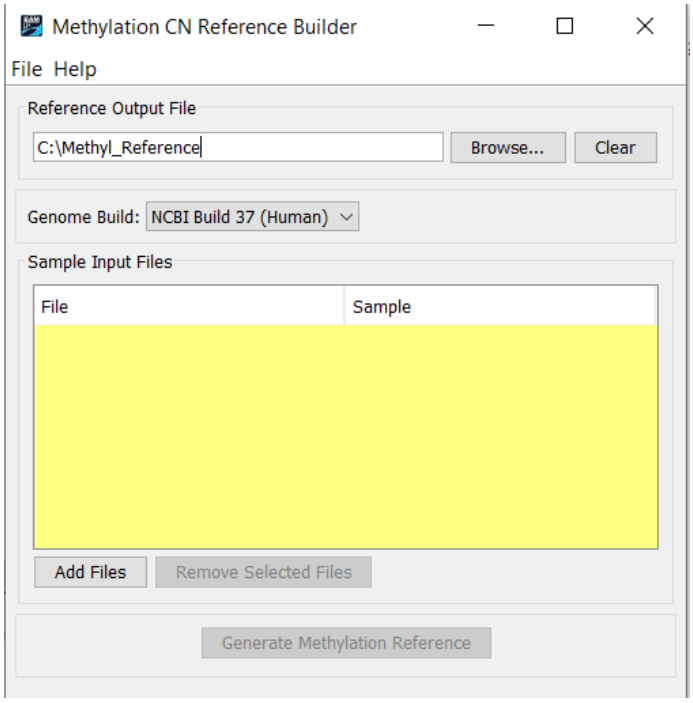


Figure 2317. User interface (UI) for the Methylation CN Reference Builder.

The program will produce a Methylation Reference File (.MRF) in the designated location, which should then be imported into VIA. While logged into the VIA client with admin privileges, navigate to the **Methylation References** tab. Select the **+** button to navigate to the desired MRF file to load into the software available for subsequent sample processing.

METHYLATION SAMPLE TYPES

Create a sample type under the Methylation sample class. There are two options available for the experimental data type: a final report exported from a methylation project can be used, or the intensity scan data (IDAT) can be used directly from the scan folder. The appropriate data type and assay name must be designated for the sample type, as shown in **Figure 238**.

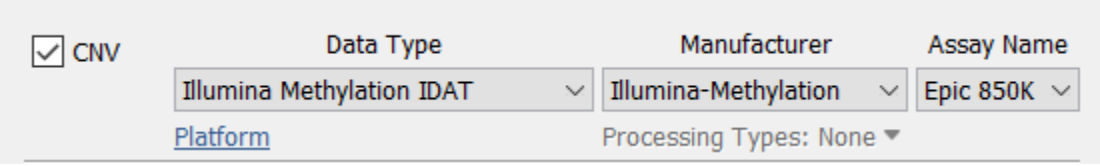


Figure 2328. Creating a sample type.

Refer to the section in this document on creating and activating processing settings. The FASST2, FASST3, and Rank segmentation algorithms are available for configuration as BAF data is not extracted from methylation data. The appropriate Reference file (MRF) needs to be designated during sample import, shown in **Figure 239**. An example of a specific EPIC Methylation sample type in shown in **Figure 240**.

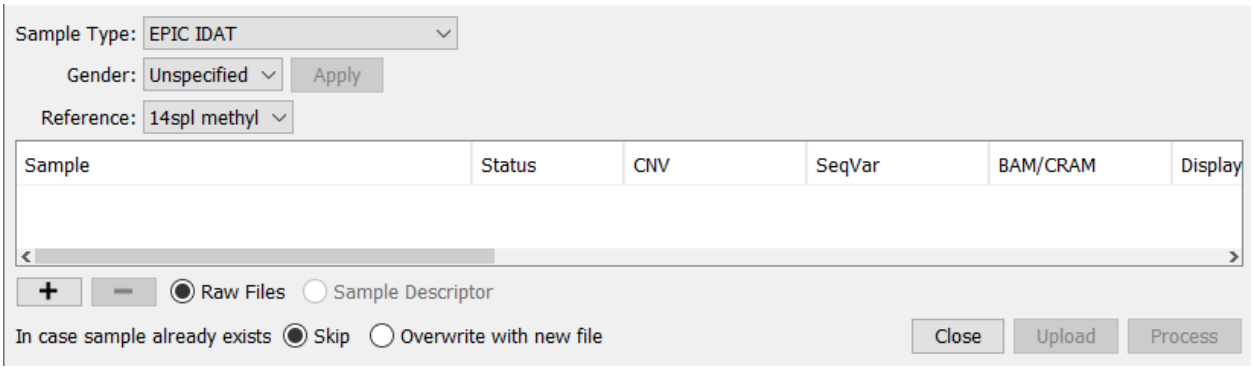


Figure 2339. MRF designation.

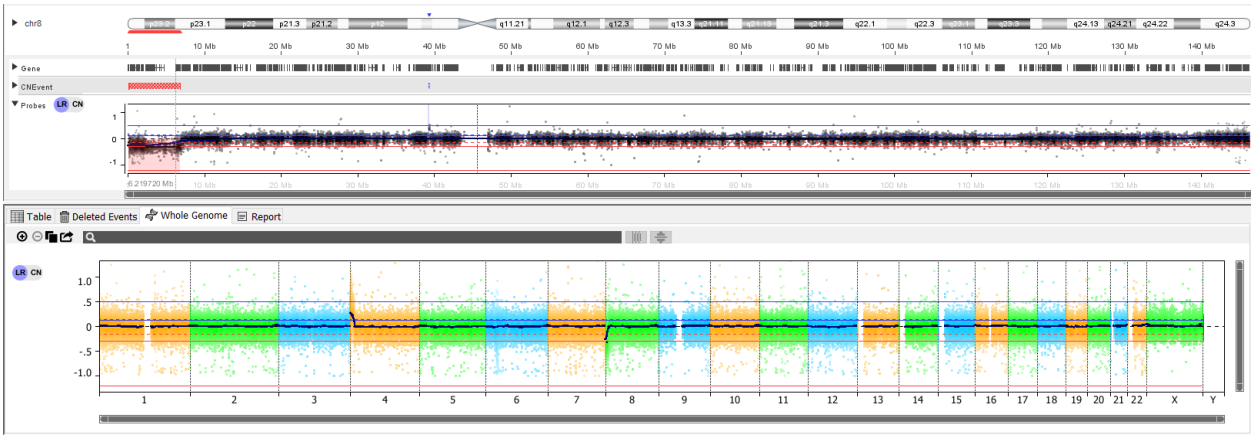


Figure 23440. Example of EPIC Methylation data display in VIA.

Homologous Recombination Deficiency Analysis

Genomic Instability Scoring for HRD

Homologous recombination deficiency (HRD) is the inability to repair double-stranded DNA breaks using the HRR cellular pathway, which consequentially results in an acquired chromosomal breakage. Clinical research has shown that cells with HRD are more sensitive to certain therapies and a measurement of HRD can be an effective pharmacogenetic biomarker across various tumor types. To provide a functional evaluation of HR status, HRD genomic scarring is an analysis approach to assess three specific quantifiable signatures of HRD genomic instability. VIA includes a measurement of these three genomic scars to aid with HRD status assessment in cancer samples across technology types.

HRD Genomic Scar Processing and Definitions

Within the Admin section for sample types that are set to an Oncology Test Type, selecting the automated Perform Genomic Scar Calculation checkbox will activate the analysis during the processing for all associated samples. The *VIA Theory of Operations* document can be referenced for a detailed description of the genomic scar measurement process in VIA and can be obtained by contacting software-support@bionano.com. In brief, genomic CNV and AOH profiles generated across data types are analyzed for scar characteristics through the implementation of three processing steps:

- Merging the CN Event and Zygosity tracks (used for HRD-LOH, TAI and LST)
- Smoothing the resulting merged track to combine similar event types and across small gaps as well as the centromere (performed independently for each scar based on parameters).
- Selecting the resulting calls or breakpoints that comply with each scar's specifications (performed independently for each scar).

The applied definition of each scar is:

- **Percent of genomic loss of heterozygosity (%gLOH)** – ratio of autosomal mono-allelic events (from both AOH and losses) as defined in the Platforms settings for "minimum LOH length" and "min. number of probes per segment", respectively.
- **Loss of heterozygosity (HRD-LOH)** - number of regions representing one parental allele resulting from a copy number neutral, or a loss, event that is longer than a specified minimum LOH event size, but shorter than the whole chromosome.
- **Telomeric Allelic Imbalance (TAI)** - number of regions with CNV or allelic imbalance longer than the specified minimum TAI event that extends to one of the telomeres but does not cross the centromere.
- **Large-Scale State Transitions (LST)** - number of chromosomal break points between adjacent regions of change in copy number or allelic content longer than a specified minimum LST event size. Adjacent events with a gap less than the maximum LST gap size are merged. State changes at centromeres and telomeres are excluded.

The characteristic event size and gap size for each genomic scar is configurable. A config file HRD Parameters is retained as a TXT file within the VIA server (../VIA Server/Storage/Resources) that can be modified to adjust the default parameters and refine the scarring performance accordingly.

The specific parameters used in calculation of the genomic scars are the minimum event size and the maximum gap size for all three scar types (HRD-LOH, TAI and LST).

HRD Genomic Scar Analysis and Display

Genomic scar measurements are conducted automatically during sample processing based on the presented CNV and AOH profiles detected by the applied processing settings. The genomic scar selection is dynamic to account for the manual changes made by an analyst's expert review and modification of the sample's CNV and AOH event calls. The Genomic scar analysis is displayed in multiple sections of the software to provide transparency and clarity to the tumor profile. The tally of each genomic scar is displayed on the **Home** page for the sample.

A display of genomic scar values including a breakdown of corresponding breakpoints/genomic regions is provided in the Sample Information window and listed in the table export, as seen in **Figure 241**.

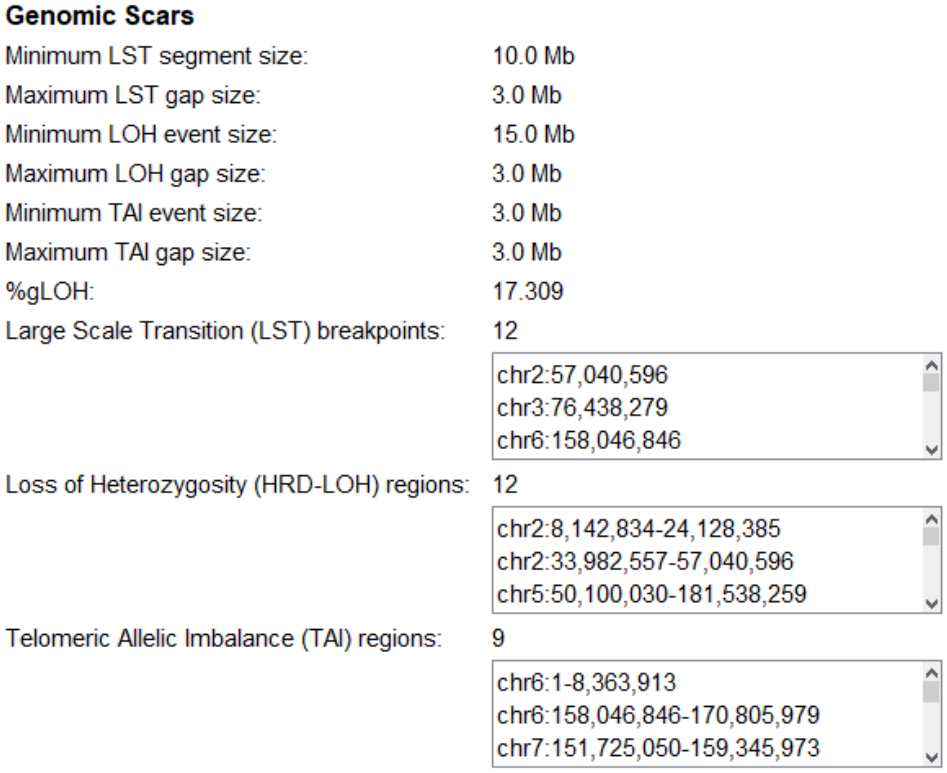


Figure 23541. Sample Information window displaying genomic scar scores.

A visual representation of the genomic scar measures plotted on chromosome ideograms in the **Genomic Scars** tab with the ability to view the scars alone or including CNV and AOH events is displayed in **Figure 242** and **Figure 243**.

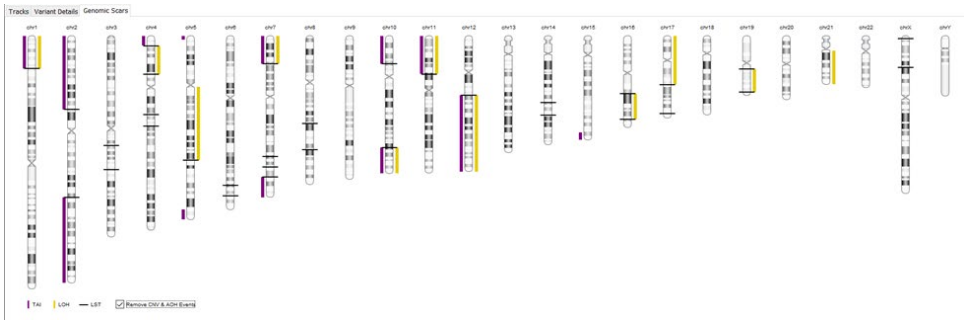


Figure 23642. Screen image visualizing genomic scars without CNV and AOH events displayed.

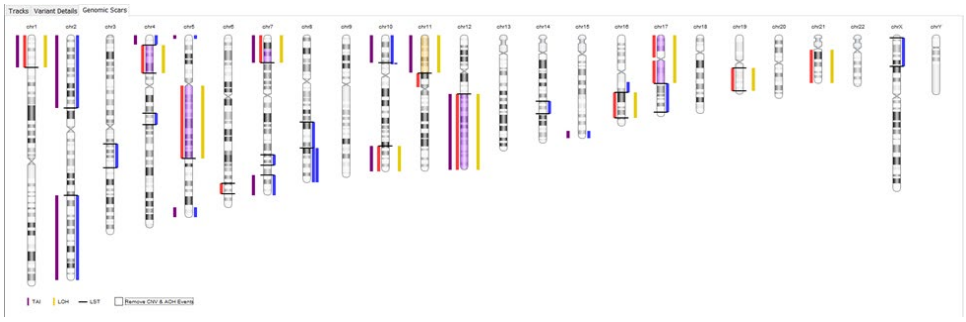


Figure 23743. Screen image visualizing genomic scars with CNV and AOH events displayed.

American College of Medical Genetics Scoreboard

The American College of Medical Genetics, or ACMG, Scoreboard feature is visible if:

- The software being used is Version 6.1 or higher.
- The test type is defined as Constitutional.
- New Dosage Sensitivity tracks (ClinGen Dosage Sensitive Map Haploinsufficiency Canonical Transcript and ClinGen Dosage Sensitive Map Haploinsufficiency Gene Components) must be present.

ACMG published updated technical standards for the interpretation and reporting of constitutional CNVs in 2019 and the Scoreboard can be used to automatically calculate the scores for many of the evidence categories described by the ACMG standards. This Scoreboard is displayed at the bottom of the **Variant Details** tab. Scores for each of the evidence categories are auto-filled wherever possible and added together resulting in the final overall score and subsequent classification of the event, as shown in **Figure 244**.

The ACMG Scoreboard is designed so users can manually add additional evidence score(s) and modify them for each evidence category using professional expertise to arrive at a final interpretation. By clicking on the **Update** button, the detailed evidence criteria are displayed with the ability to modify scores for any evidence criteria. Additionally, a **Notes** section for each evidence category is provided so the user can input text with regards to why that score was entered.

When the fields in the 2019 CN Guidelines Scoreboard are shaded yellow, that implies that no manual updates were made to any of the scores. Upon manual editing of any of the evidence criteria, the fields on the Scoreboard are shaded white.

Overall, the Scoreboard provides a formal reasoning structure that standardizes CNV event classification and ultimately helps reduce errors by eliminating the need to use external tools. An example is provided in **Figure 245**.

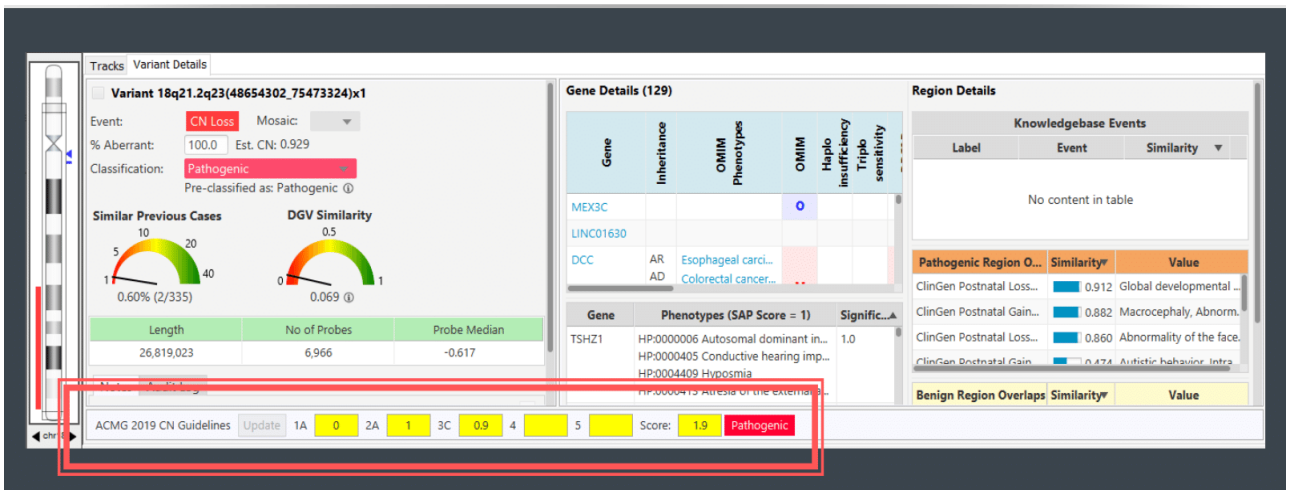
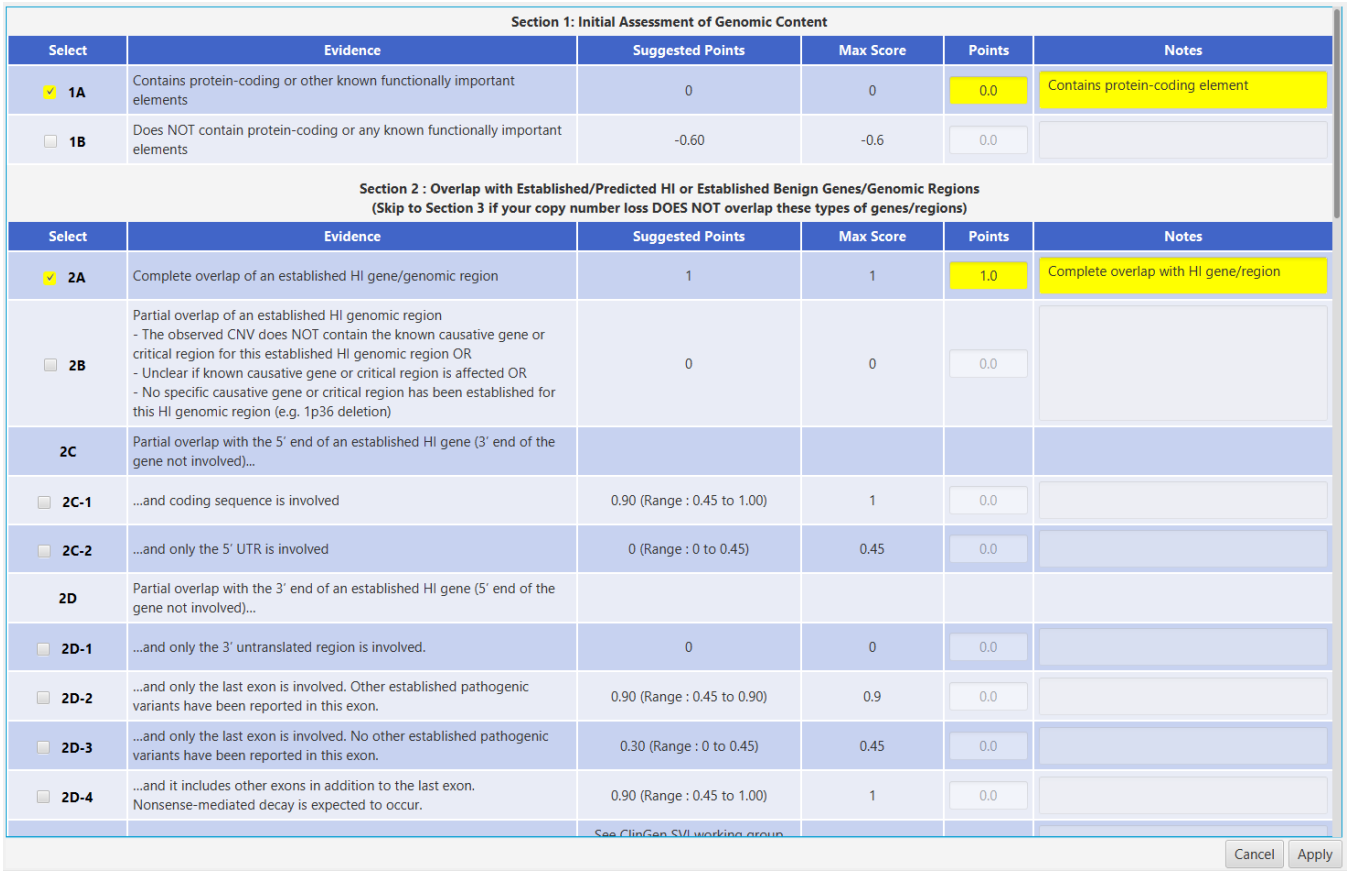


Figure 2384. ACMG Guidelines.



Section 1: Initial Assessment of Genomic Content

Select	Evidence	Suggested Points	Max Score	Points	Notes
1A	Contains protein-coding or other known functionally important elements	0	0	0.0	Contains protein-coding element
1B	Does NOT contain protein-coding or any known functionally important elements	-0.60	-0.6	0.0	

Section 2 : Overlap with Established/Predicted HI or Established Benign Genes/Genomic Regions
(Skip to Section 3 if your copy number loss DOES NOT overlap these types of genes/regions)

Select	Evidence	Suggested Points	Max Score	Points	Notes
2A	Complete overlap of an established HI gene/genomic region	1	1	1.0	Complete overlap with HI gene/region
2B	Partial overlap of an established HI genomic region - The observed CNV does NOT contain the known causative gene or critical region for this established HI genomic region OR - Unclear if known causative gene or critical region is affected OR - No specific causative gene or critical region has been established for this HI genomic region (e.g. 1p36 deletion)	0	0	0.0	
2C	Partial overlap with the 5′ end of an established HI gene (3′ end of the gene not involved)...				
2C-1	...and coding sequence is involved	0.90 (Range : 0.45 to 1.00)	1	0.0	
2C-2	...and only the 5′ UTR is involved	0 (Range : 0 to 0.45)	0.45	0.0	
2D	Partial overlap with the 3′ end of an established HI gene (5′ end of the gene not involved)...				
2D-1	...and only the 3′ untranslated region is involved.	0	0	0.0	
2D-2	...and only the last exon is involved. Other established pathogenic variants have been reported in this exon.	0.90 (Range : 0.45 to 0.90)	0.9	0.0	
2D-3	...and only the last exon is involved. No other established pathogenic variants have been reported in this exon.	0.30 (Range : 0 to 0.45)	0.45	0.0	
2D-4	...and it includes other exons in addition to the last exon. Nonsense-mediated decay is expected to occur.	0.90 (Range : 0.45 to 1.00)	1	0.0	

Figure 2395. Partial view of the expanded **Points** and **Notes** sections.

In addition to functioning as a stand-alone guideline, the Scoreboard can be coupled with the automated variant pre-classification decision tree to pre-classify events. This allows users to quickly sort through the events on the table and prioritize for review. See **Figure 246** for an illustration of this feature.

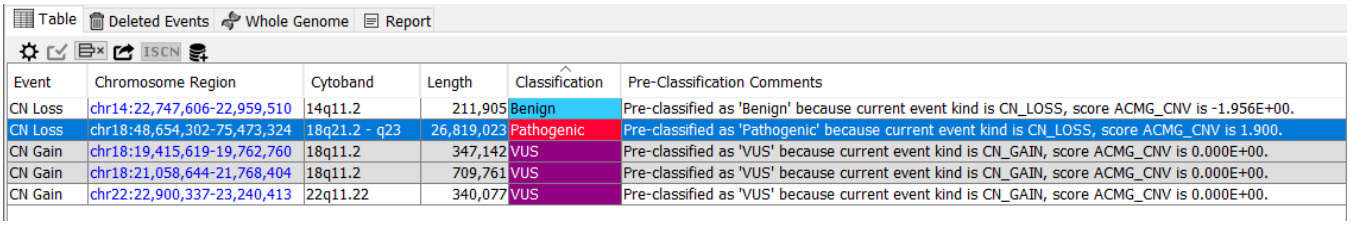


Table | Deleted Events | Whole Genome | Report

Event	Chromosome Region	Cytoband	Length	Classification	Pre-Classification Comments
CN Loss	chr14:22,747,606-22,959,510	14q11.2	211,905	Benign	Pre-classified as 'Benign' because current event kind is CN_LOSS, score ACMG_CNV is -1.956E+00.
CN Loss	chr18:48,654,302-75,473,324	18q21.2 - q23	26,819,023	Pathogenic	Pre-classified as 'Pathogenic' because current event kind is CN_LOSS, score ACMG_CNV is 1.900.
CN Gain	chr18:19,415,619-19,762,760	18q11.2	347,142	VUS	Pre-classified as 'VUS' because current event kind is CN_GAIN, score ACMG_CNV is 0.000E+00.
CN Gain	chr18:21,058,644-21,768,404	18q11.2	709,761	VUS	Pre-classified as 'VUS' because current event kind is CN_GAIN, score ACMG_CNV is 0.000E+00.
CN Gain	chr22:22,900,337-23,240,413	22q11.22	340,077	VUS	Pre-classified as 'VUS' because current event kind is CN_GAIN, score ACMG_CNV is 0.000E+00.

Figure 2406. Events table sorted by the classification column for pre-classified events using a decision tree.

Using a Decision Tree: An Overview

How to Increase Efficiency and Turnaround

A decision tree (DT) is a set of logic rules used to automatically pre-classify events in the sample. The use of a decision tree can help increase the efficiency of the sample review process. High-throughput technologies create too much information for an individual reviewer to analyze manually while maintaining efficiency. Many events (e.g., CNVs, AOH) can be detected in a sample but the goal is to narrow down the number of events to a manageable quantity. The automated classification system in VIA increases efficiency by automating much of the process of classifying events. A reviewer's objective is to classify the identified events and label them (e.g., pathogenic, likely pathogenic, benign), and many of these events can be pre-classified based on a set of logic rules added to the system and defined by the Administrator. The pre-classified events are labeled as such and provide the reasoning that marked them with that classification. The approach uses external databases (e.g., ClinGen, OMIM, DECIPHER) as well as internal data generated over time in the local database. The Administrator creates logic rules for automatic decision tree pre-classification (see section on Pre-classification Syntax). A DT is associated with a specific sample type, and when a sample is loaded, the DT logic runs and pre-classifies events based on the rules. The event pre-classification results can be reviewed in the Classification column in the table of results in the **Sample Review** window (**Figure 247**).

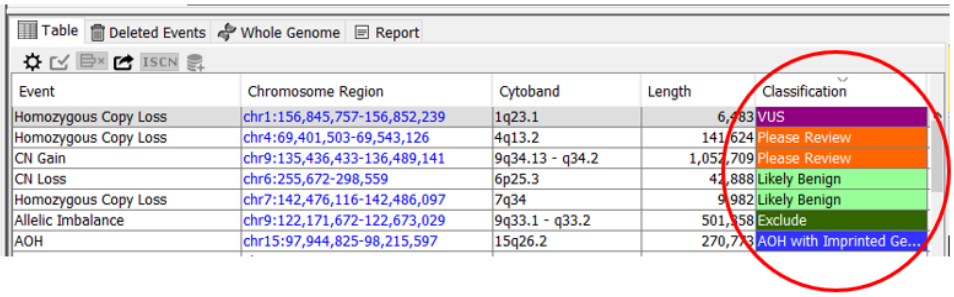


Figure 2417. Example of pre-classified events viewable in the **Results** table in the **Classification** column.

During VIA installation, Bionano Support can assist with configuration and in generating the DT language and scripts when setting up sample types. Support will generate scripts to mirror the logic used in the lab for the interpretation process and suit the DT to various analysis workflows, such as constitutional and oncology.

The Administrator can create multiple decision trees for a single sample type to test them and then choose one to associate with a sample type. Through the Administrator interface, all decision trees defined for a sample type are run when a sample is processed. The results can be displayed in the **Results** table, one in each column for each DT (**Figure 248**). This allows the Administrator to see how each DT is performing and then select one to associate with a sample type for future processing performed by other users. Only when logged in as Administrator can all decision trees for a sample type be added and run during sample processing.

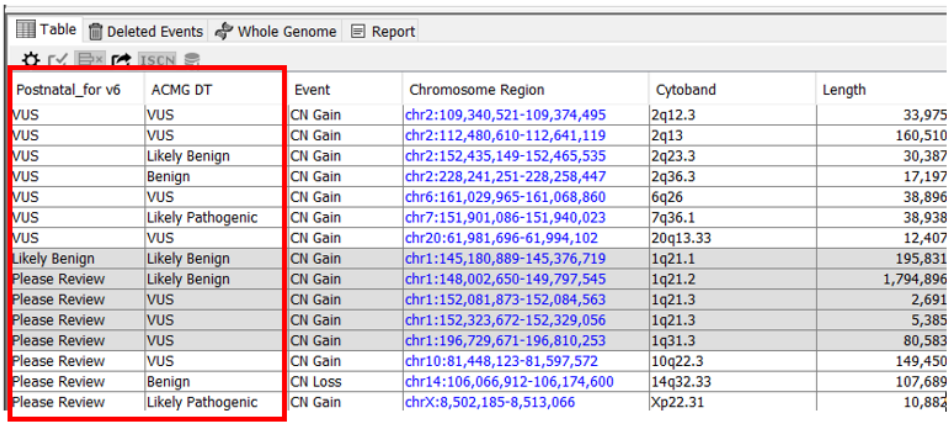


Table | Deleted Events | Whole Genome | Report

Postnatal_for v6	ACMG DT	Event	Chromosome Region	Cytoband	Length
VUS	VUS	CN Gain	chr2:109,340,521-109,374,495	2q12.3	33,975
VUS	VUS	CN Gain	chr2:112,480,610-112,641,119	2q13	160,510
VUS	Likely Benign	CN Gain	chr2:152,435,149-152,465,535	2q23.3	30,387
VUS	Benign	CN Gain	chr2:228,241,251-228,258,447	2q36.3	17,197
VUS	VUS	CN Gain	chr6:161,029,965-161,068,860	6q26	38,896
VUS	Likely Pathogenic	CN Gain	chr7:151,901,086-151,940,023	7q36.1	38,938
VUS	VUS	CN Gain	chr20:61,981,696-61,994,102	20q13.33	12,407
Likely Benign	Likely Benign	CN Gain	chr1:145,180,889-145,376,719	1q21.1	195,831
Please Review	Likely Benign	CN Gain	chr1:148,002,650-149,797,545	1q21.2	1,794,896
Please Review	VUS	CN Gain	chr1:152,081,873-152,084,563	1q21.3	2,691
Please Review	VUS	CN Gain	chr1:152,323,672-152,329,056	1q21.3	5,385
Please Review	VUS	CN Gain	chr1:196,729,671-196,810,253	1q31.3	80,583
Please Review	VUS	CN Gain	chr10:81,448,123-81,597,572	10q22.3	149,450
Please Review	Benign	CN Loss	chr14:106,066,912-106,174,600	14q32.33	107,689
Please Review	Likely Pathogenic	CN Gain	chrX:8,502,185-8,513,066	Xp22.31	10,882

Figure 2428. Sample types with multiple DTs: each DT is run, and the results can be visualized using the Administrator log in.

Select the desired sample type and add the DT script in the **Decision Trees** tab by clicking the **Add** (**+**) tool (**Figure 249**). In the **Event Classification** tab, ensure all the different event classification values that are present in the DT script are added. For each classification value, select a desired color for graphical representation of that classification. Use the **Add**, **Remove**, and **Edit** tools to add classification values and to make changes to the available classification values. The DT to apply for automated pre-classification can be selected using the dropdown field in the **Event Classification** tab (red arrow in **Figure 250**). Classified events displayed in the tracks and table are color coded based on the selection for each classification value in the **Event Classification** tab. When any changes are made to events in a sample the DT script can be run again if the option **Enable Automated Pre-classification on manual edits** is selected (circled in **Figure 250**). When an event is manually altered (e.g., event boundaries were changed), a pop-up alert will ask the user if automated pre-classification should be run, and the user is given a **Yes** or **No** option. Classified results are shown as in **Figure 251**.

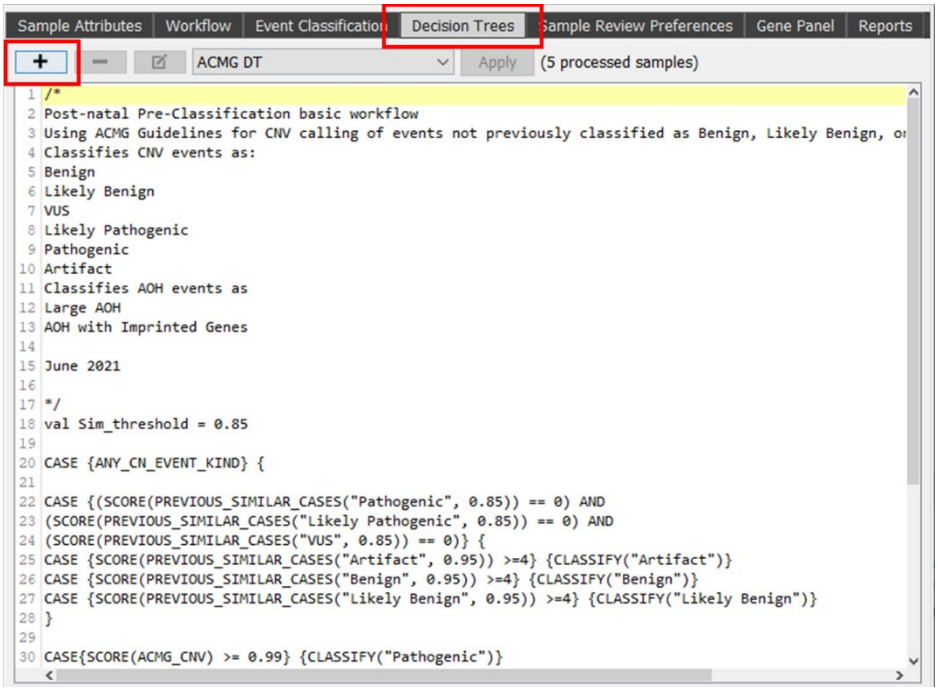


Figure 2439. The DT script is added to the Sample Type.

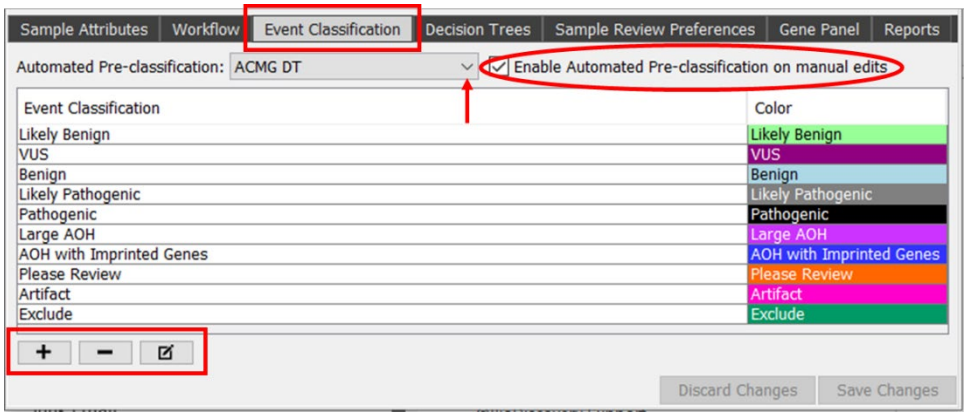


Figure 24450. Graphical representation of each classification value in the **Event Classification** tab.

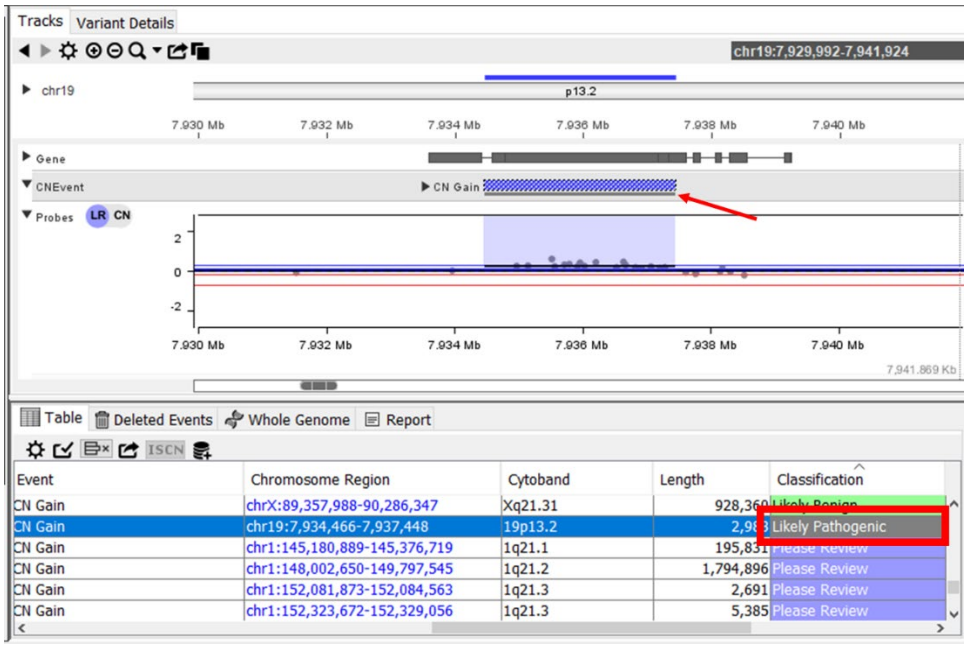


Figure 24551. Classified events displayed in the **Results** table.

Filtering of CNV, Allelic Events, and Sequence Variant Data

An initial filtering schema is needed to narrow down the list of events to display in the VIA browser and table. Events are removed/included from user view based on options selected using the **Filters** and **Panels** tabs. These tabs are located to the right of the table at the bottom of the window, as seen in **Figure 252**. If there are no **Filters** or **Panels** applied, the tabs will be gray/white, as shown in the image below. If filters or panels **ARE** applied, the tab will be gold.

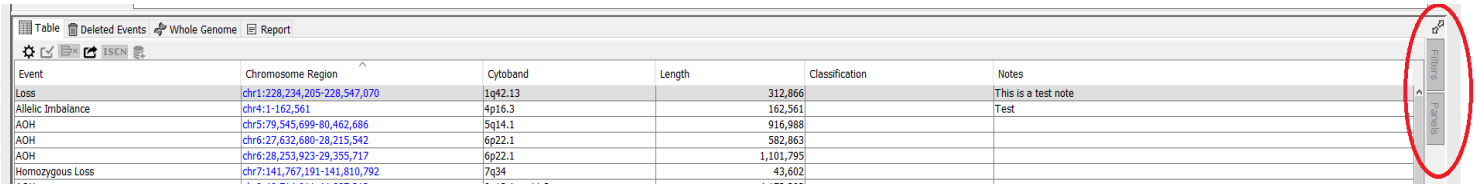


Figure 24652. Filters and **Panels** tabs.

When open, the width of the **Filters** and **Panels** section can be adjusted to become wider or narrower by hovering over the left boundary until a horizontal double-sided arrow appears and then clicking and dragging the left boundary.

Clicking on the **Filters** tab will display the filtering options. Depending on the sample type, various additional tabs may be displayed in the panel. Filters are organized by modality and separated into four tabs: **Copy Number**, **Allelic**, **Sequence Variants**, and **SV Events**. **Figure 253** below displays all four tabs as the example sample has all four types of variants.

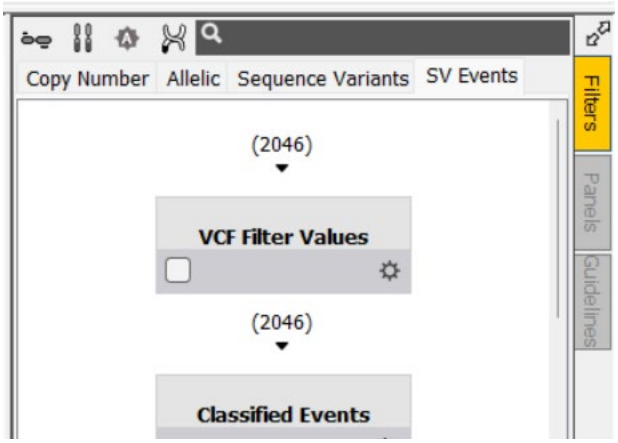


Figure 24753. Four variants.

Each tab displays boxes with details for each filter and icons indicating the status of that filter. Some filters are specific to the data modality and others apply across multiple modalities and will be displayed in each tab. **Gear** icons at bottom right indicate that the filter has parameters needed for selection as well as whether the filter affects other filters. Clicking on the icon opens the filter parameters in another window.

Gear icons with arrows indicate that this filter is linked across tabs; in other words, the same filter appears in other tabs. Parameters selected in one tab are carried to all the other tabs. Also, if the filter is applied in one tab, it is automatically applied in other tabs as well. For example, if, in the **Sequence Variants** tab, the **Classified Events** filter is enabled, it will automatically be enabled in the **Copy Number** and **Allelic** tabs as well (if those tabs exist).

The checkbox on the bottom left is for turning the filter on and off (checked/unchecked, respectively) but is not displayed unless a filter parameter has been selected. For example, the checkbox will not be visible because a panel is not selected in the filtering parameters. Note that the selected gene panel name will also be displayed in the filter box.

Some filters will have an arrow next to the checkbox. An arrow pointing down indicates that variants after this stage are sent to a filter in a different tab and input to that filter. An arrow pointing up indicates that variants coming into this filter are coming from a specific stage of a filter in another tab.

The panel filter displays the parameter directly in the filter box. This is for those filters where there is only one option to select and then this parameter is displayed in the box. In the example above, after selecting a panel gene list, the box displays the panel selected (solid tumor panel).

To enable a filter, mark the checkbox. In **Figure 254**, one box is not checked, and one can see the number of variants below the box, which is 2917. Checking off the box enables the filter, and the number of variants decreases as some are filtered out due to the panel selection.

Figure 2484. Decrease in variants.

Visibility of different filters is also dependent on the information available for the Sample. If there are no linked family relationships for a sample, for example, the Inheritance Pattern filter will not be displayed in the filter chain.

The Classified Events filter provides several options for treating each classification category. The choices are to Always Show, Do Nothing, and Remove; these options are available for all filters. **Figure 255** displays the parameters for all filter classifications and explanations per radio button are provided below.

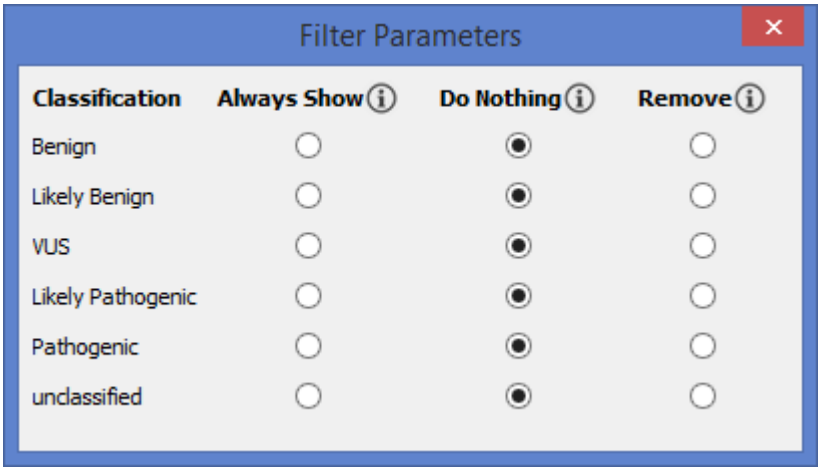


Figure 2495. Radio buttons for treatment and classification.

Always Show: Selecting this button will always display events classified as such, regardless of the filters applied downstream. For example, if Benign is marked as **Always Show**, then events classified as Benign will be displayed in the **Browser and Results** table even if downstream events are filtered out. If there are a total of 392 events prior to application of this filter and there are two Benign events in the results, then selecting **Always Show** for Benign will still display 392 events in the **Results** table and the counts at the bottom of the table.

Do nothing: Selecting this radio button does not apply the filter to this classification; in other words, this selection will pass all variants of this classification to the next filter in the chain.

Remove: Selecting this radio button will remove the event from the browser and table display. If Benign is marked as Remove, then Benign events will not be displayed in the browser or table. If Benign is marked as Remove and there are a total of twenty-five events prior to this filter application but one event is Benign, after enabling this filter, the benign event will be removed and the total number of events at the bottom of the table will show 24.

The **Panel Selection** filter shows only events in the gene panel list. If at least one gene panel is associated with the sample type, this filter will be present in the filter chain. All available panels will be listed, and the user can select one from the dropdown menu. **Figure 256** depicts the UI for Panel Selection.

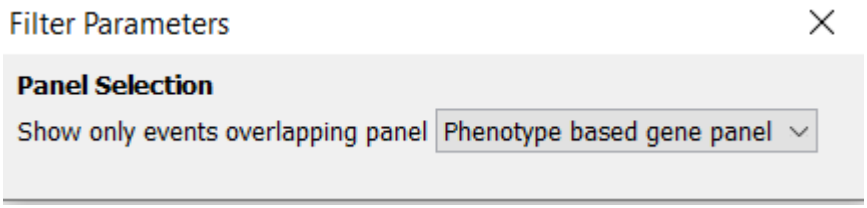


Figure 2506. Dropdown menu for **Panel Selection.**

Similar previous cases are events that appear repeatedly in the database by percentage or by number of cases. Marking relevant checkboxes to remove unwanted events depending on how often a similar event is detected in the VIA database is displayed below in the UI image and in **Figure 257**.

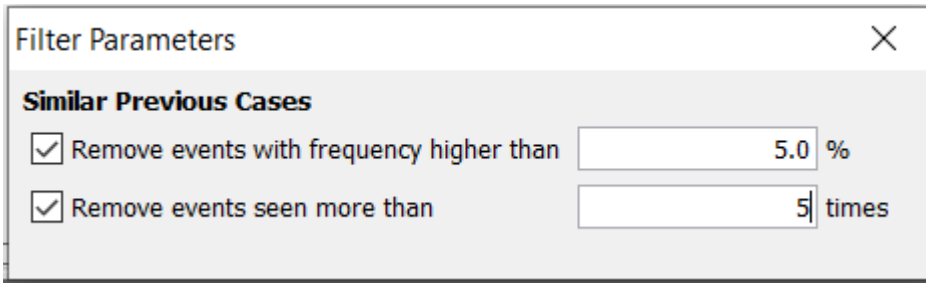


Figure 2517. Like previous filter selection.

The **Event Types** filter allows the user to mark checkboxes to remove certain event types as depicted in the UI image and in **Figure 258**, much in the same manner as previously described cases.

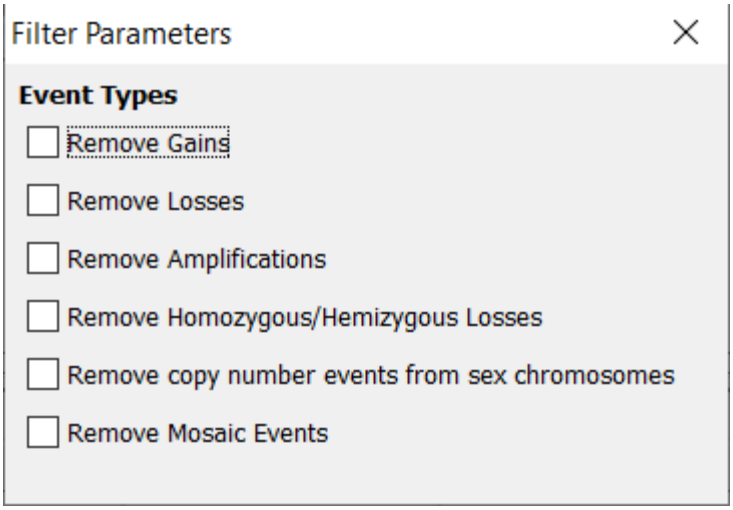


Figure 2528. Event Types filter.

Users can mark relevant checkboxes and specify a size or number of probes to remove those events, as shown in **Figure 259**.

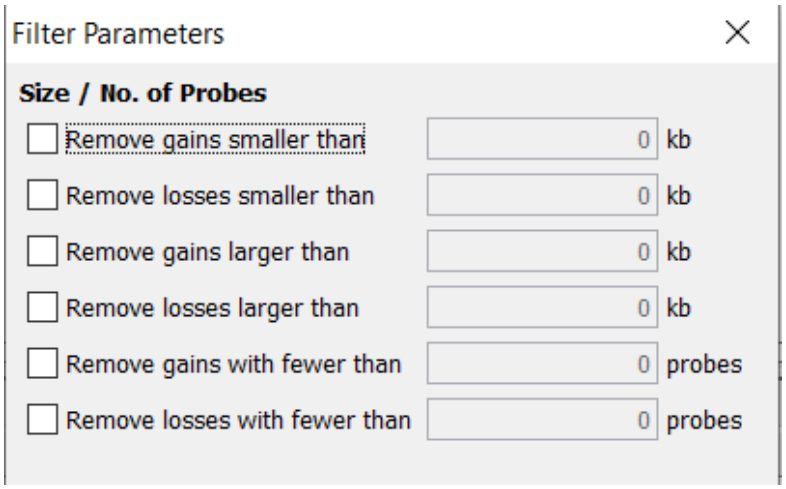


Figure 2539. Size/number of probes.

Mark relevant checkboxes to remove items not covered by an allelic event, as shown in **Figure 260**.

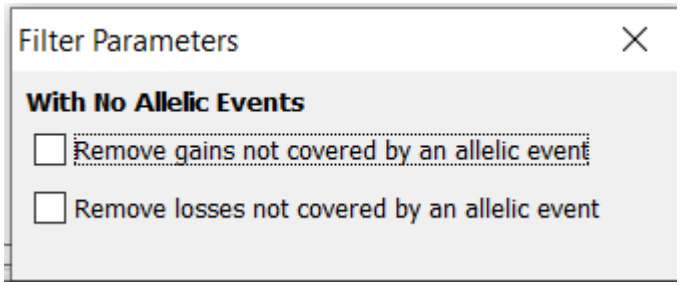


Figure 25460. With no allelic events.

Allelic Event Filters

The Classified events filter is shared with CNV events. The **Panel Selection** and **Similar to Previous** filters are like the CNV events but can be configured separately for allelic events.

Mark relevant checkboxes to remove events that come up repeatedly in the database (by percentage or by number of cases). See **Figure 261** and **Figure 262**.

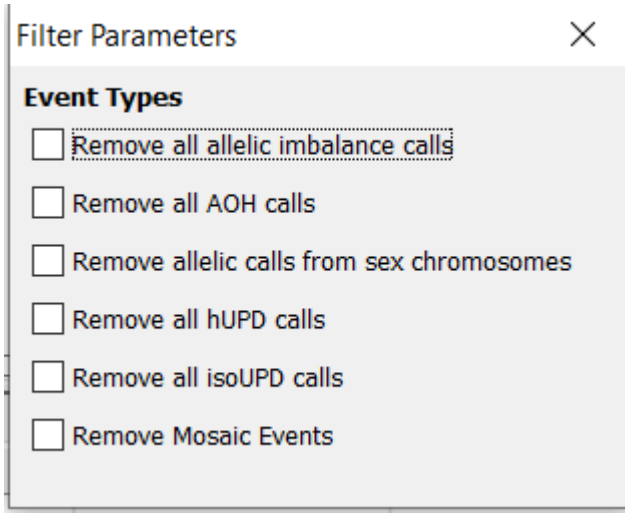


Figure 2551. Event types.

Mark relevant checkboxes and specify a size or number of probes to remove those events, as shown in **Figure 262**.

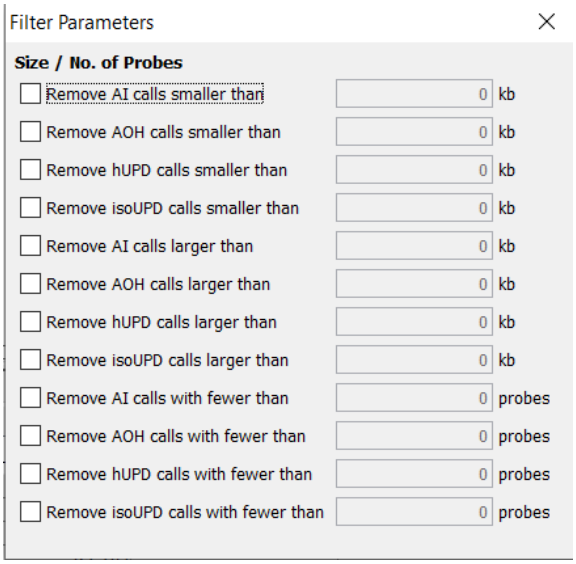


Figure 2562. Remove allelic events based on size or number of probes.

Mark relevant checkboxes to remove events not covered by copy number event, as seen in **Figure 263**.

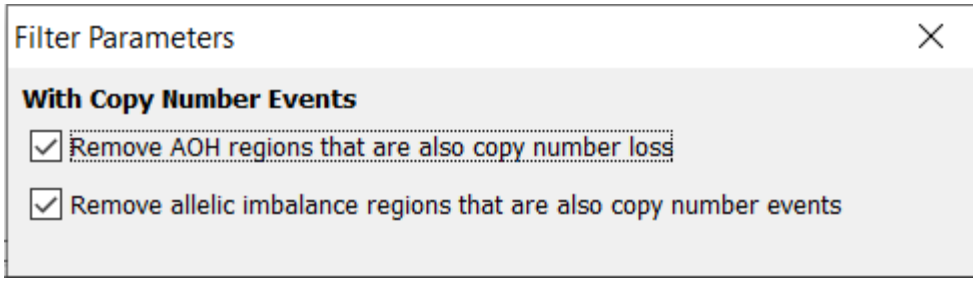


Figure 2573. Remove events not covered by copy number.

Mark the checkbox to show only AOH and/or allelic imbalance events that overlap SeqVar (Sequence Variant) events, as seen in **Figure 264**. This includes SeqVar events after the Interest Level stage in the **Sequence Variants** tab.

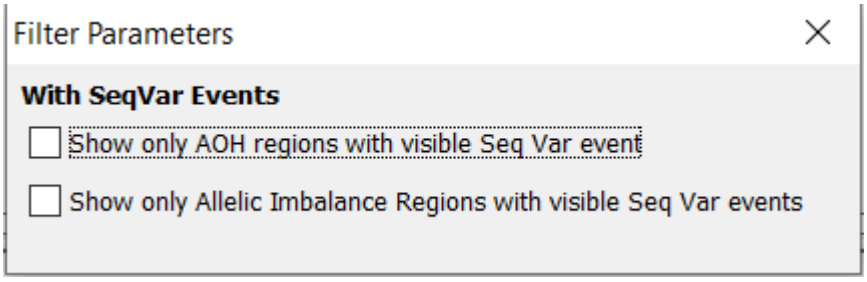


Figure 2584. SeqVar overlap.

Sequence Variant Filters

CLINVAR CLASSIFICATION

The Sequence Variant filter provides several options (see **Figure 265**) to classify events based on the ClinVar Classification and ClinVar Star Rating System. The Minimum Rating is a drop down that can be changed between 1, 2, 3 or 4 stars.

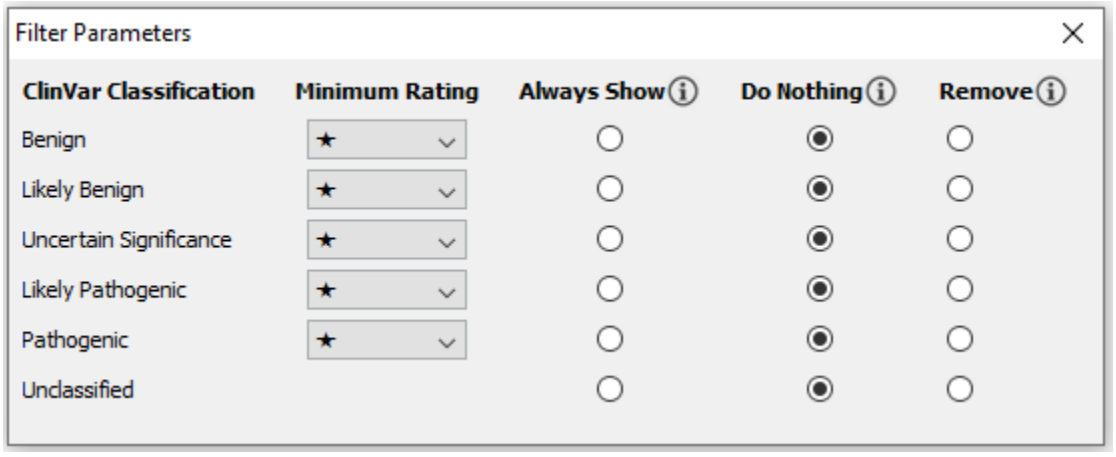


Figure 2595. Minimum Rating menu.

ClinVar uses the star system as described below in **Figure 266** and can be obtained from the National Institutes of Health website.

NOTE: Events not classified by ClinVar will not be filtered.

Number of gold stars	Review status	Description
four	practice guideline	practice guideline
three	reviewed by expert panel	reviewed by expert panel
two	criteria provided, multiple submitters, no conflicts	Two or more submitters with assertion criteria and evidence (or a public contact) provided the same interpretation.
one	criteria provided, conflicting interpretations	Multiple submitters provided assertion criteria and evidence (or a public contact) but there are conflicting interpretations. The independent values are enumerated for clinical significance.
one	criteria provided, single submitter	One submitter provided an interpretation with assertion criteria and evidence (or a public contact).
none	no assertion for the individual variant	The allele was not interpreted directly in any submission; it was submitted to ClinVar only as a component of a haplotype or a genotype.
none	no assertion criteria provided	The allele was included in a submission with an interpretation but without assertion criteria and evidence (or a public contact).
none	no assertion provided	The allele was included in a submission that did not provide an interpretation.

Figure 2606. The ClinVar Star Rating System.

The Classified events filter is shared with CNV and Allelic Events. The **Panel Selection** and **Similar to Previous** filters are like the CNV and Allelic events but can be configured separately for sequence events. Events provide options on how to treat each classification category.

Quality and Population Frequency: Instead of filtering prior to processing (as per settings for the sample type), the population frequency, read depth, and quality of sequence variants can also be filtered dynamically within the filter chain after a sample is processed.

Populations are grouped together into a tree structure with checkboxes allowing selection of specific population filters and allow exclusion of variants based on allele frequencies for different populations from different data sources. Based on the version of annotator used, the fields available here will differ. The latest linked Nirvana Annotator uses gnomAD (Genome and Exome) as the data source. See **Figure 267**.

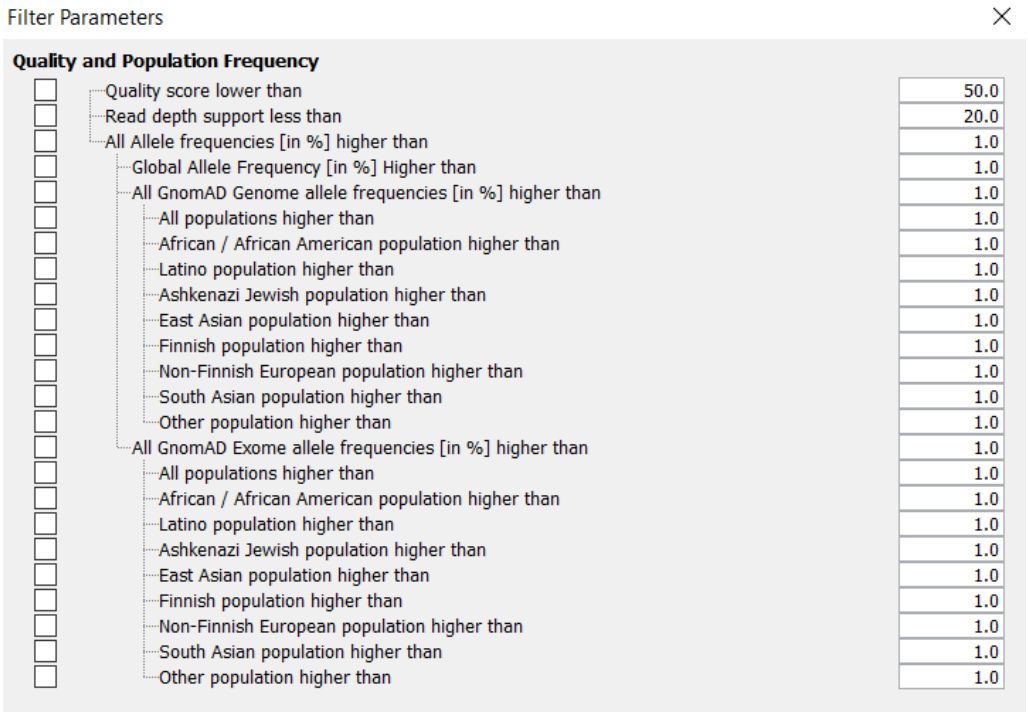


Figure 2617. Filter parameters for population frequency.

The Quality Score is the QUAL value from the VCF file, a Phred-scaled probability that a REF/ALT polymorphism exists at this site given sequencing data.

The allele frequencies are arranged in a tree-like structure such that selections in the higher nodes apply to all items under that branch. For example, to change all 1000 Genomes Project frequencies to 2.0, that change can be made in the row highlighted below and all the population specific groups' frequencies will also be at 2.0, as seen in **Figure 268**. Each individual population can have its own frequency as well and the field for that group can be edited.

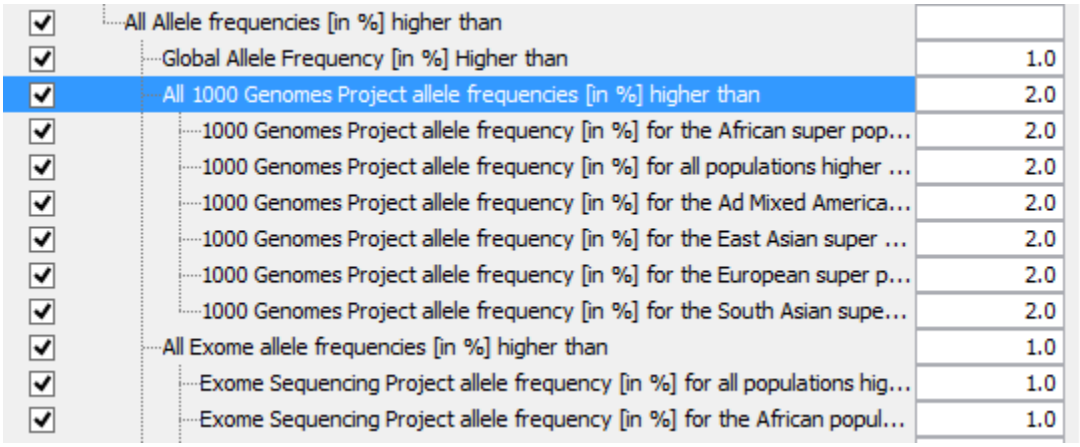


Figure 2628. Changing a frequency.

Event Types: Check boxes to display the specified events are seen in **Figure 269**. Numbers in parentheses next to the event type indicate the number of events of that type in the current sample. If none of the boxes are checked and the filter is off, all variants will be displayed (no selection is being made since the filter is not enabled).

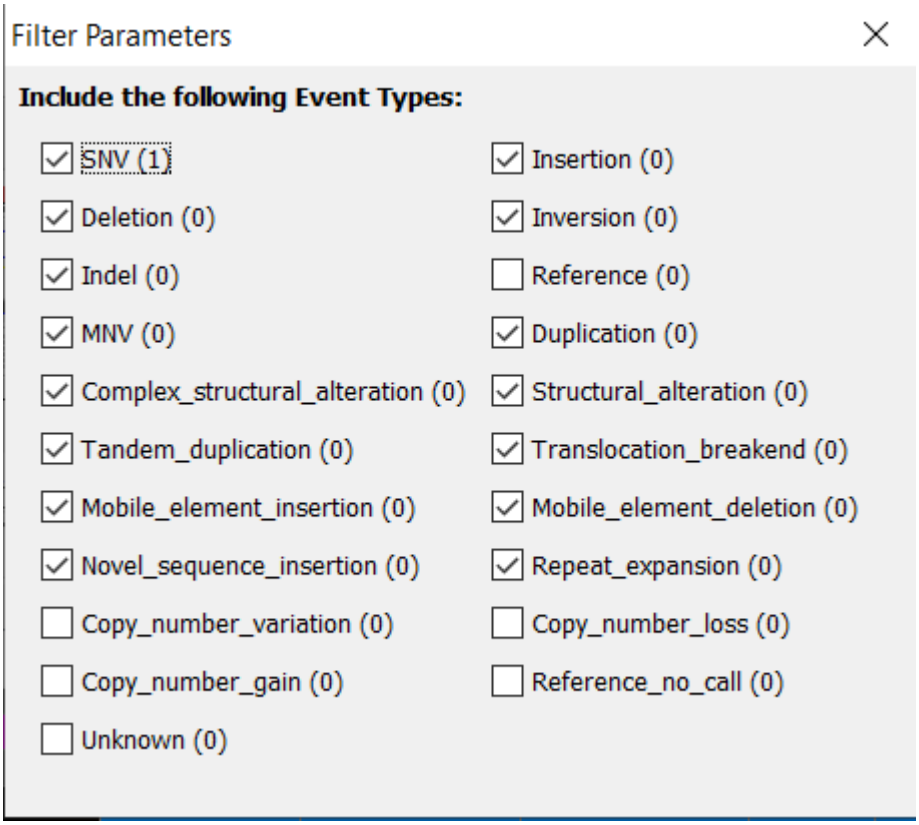


Figure 2639. Specified Event Types.

Event Consequences: Check boxes to display the specified consequences are seen in **Figure 270**. Numbers in parentheses indicate the number of events with that consequence, which are available in annotated VCF files. If the VCF is not annotated, events will not be displayed.

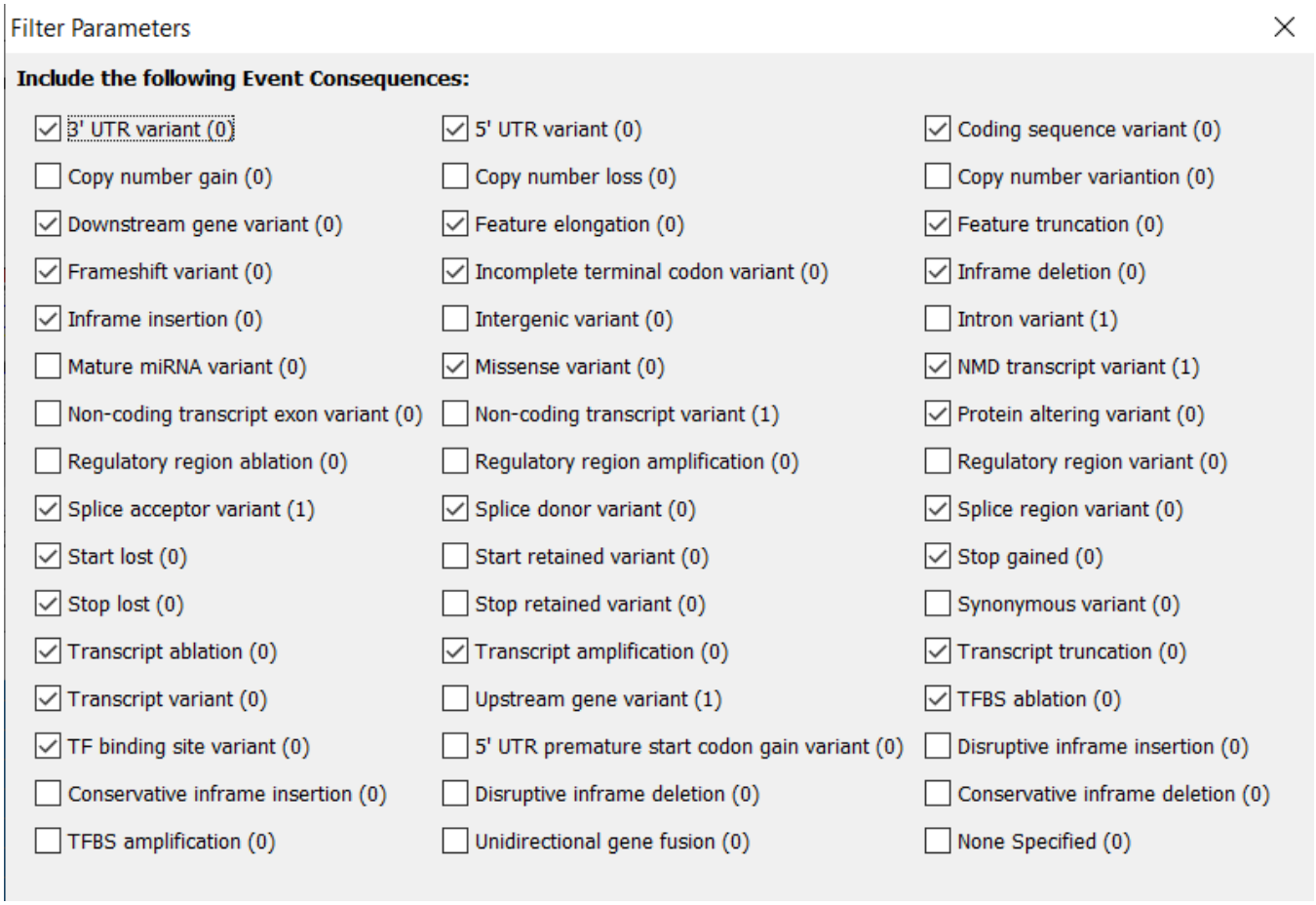


Figure 26470. Specified consequences.

If no consequences are selected and the filter is not enabled, all variants will be displayed as the filter is not on and therefore is not discriminating between different consequences. If no consequences are selected and the filter is enabled, no events will be displayed since nothing was selected to be displayed.

Zygosity Filter: The **Zygosity** filter and the **Inheritance Pattern** filter allow users to effectively filter CNV and/or SeqVar events according to genotype and particular mode of inheritance.

Zygosity filter: In the **SeqVar** filter pipeline, there are three options to filter for specific genotypes of the variant, those being heterozygous, homozygous, and hemizygous. An option to match the genotype of the variant with the mode of inheritance for the associated gene is available in this filter, shown in **Figure 271** (OMIM Inheritance Match).

NOTE: The **Zygosity** filter is disabled (grayed out) when the **Recessive Inheritance Pattern** Filter is active.

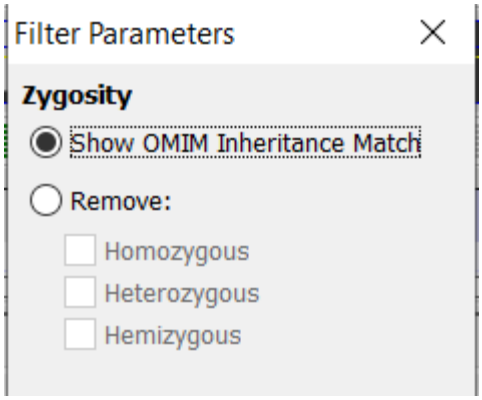


Figure 2651. Zygosity filter.

OMIM Inheritance Match: When **Show OMIM Inheritance Match** is selected, heterozygous variants in genes that are either recessive or dominant and homozygous variants in genes that are recessive are shown in the table. The OMIM dominant and recessive genes are based on the OMIM Morbid Phenotypes Dominant 5K Extended and the OMIM Morbid Phenotypes Recessive tracks that are a standard part of regions and annotations.

Variant Read Fraction and Count Filter: Check the box and edit the number in the text field to specify read fraction as a percent and actual read counts, as seen in **Figure 272**.

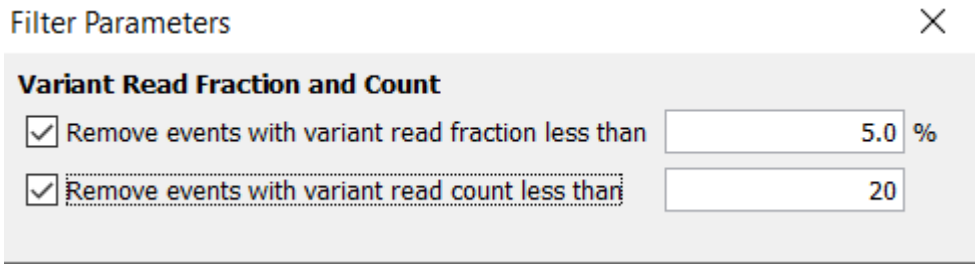


Figure 2662. Variant read fraction and count.

Interest Level: Check boxes can be used to remove all sequence variant events that are at the selected severity of the event. Variants from this stage are used to filter overlapping Allelic Events as per the With SeqVar Events filter in the **Allelic Events** tab, shown in **Figure 273**. There are two ways to select events:

1. A slider can be dragged up or down the list of interest levels to quickly select multiple items to filter out everything below the line.
2. Checkboxes next to each interest level can be selected for finer tuning of filtering to choose individual interest levels to filter out.

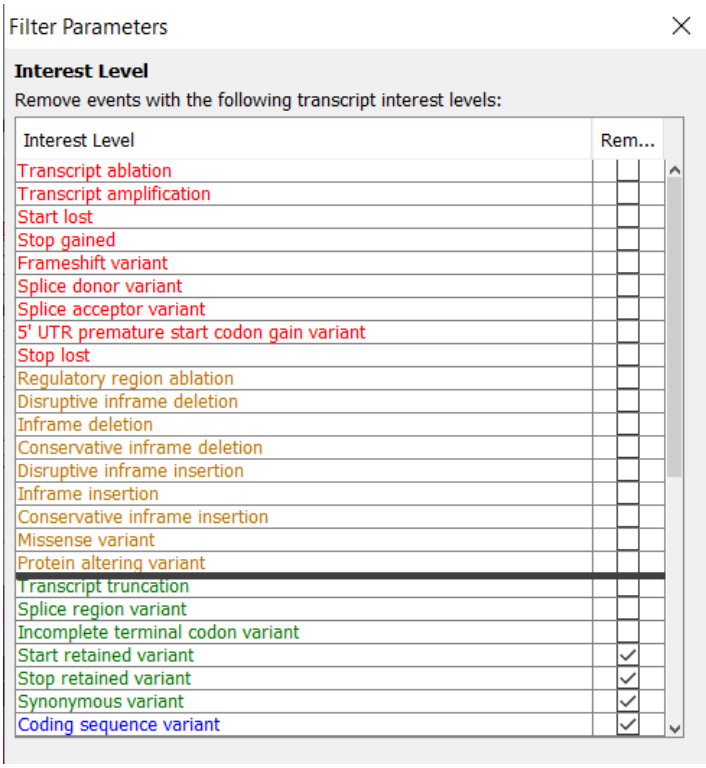


Figure 2673. Interest levels.

Values considered here are from the **Consequence** column. The interest level used for the filter is the highest one from the selected transcript.

***In Silico* Predictors:** Values considered here are from the *in silico* prediction column, shown in **Figure 274**. The highest interest level is taken from all available values in the *in silico* prediction column and that is the interest level used for the filter.

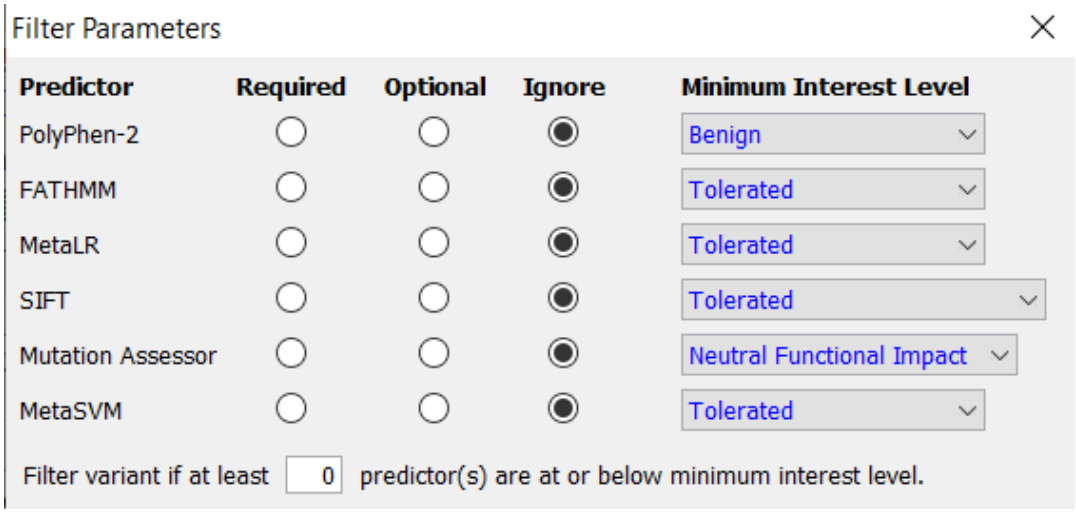


Figure 2684. *In silico* prediction column.

One or more filters may be selected and marked as Required or Optional. A minimum number of predictors that match criteria can be specified to add further assurance before filtering out variants.

Compound Events: When searching for compound events, genes may be padded by the user-specified amount to include upstream and downstream regions, shown in **Figure 275**.

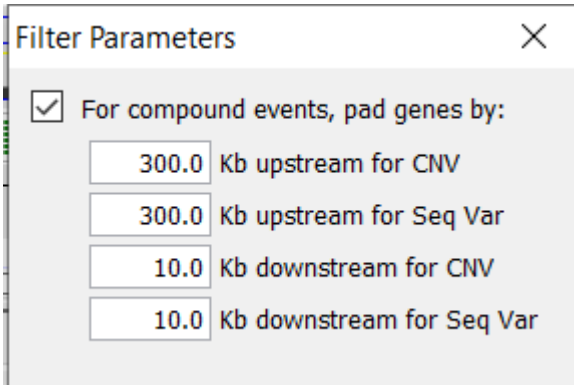


Figure 2695. Compound events filter.

Activating this filter will parse out sequence variant events that do not have another sequence variant, CN or allelic event overlapping the same gene. This filter takes in events from the end of the copy number chain and filtered sequence variants before this stage (operates only on visible sequence variant events only). For this calculation, the gene is taken in its entirety plus any padding, if specified. With copy number gain events overlapping a gene, the CN gain must have at least one breakpoint within the gene and a 30kbp padding around the gene. If the CN gain event is fully covering the gene plus padding, it is not considered in the compound event filter and the sequence variant event will not be filtered out.

Inheritance Pattern Filter: Only shows events of the selected inheritance pattern. The filter allows selection of multiple models to be applied at once and the union of these results will be displayed, as shown in **Figure 276**.

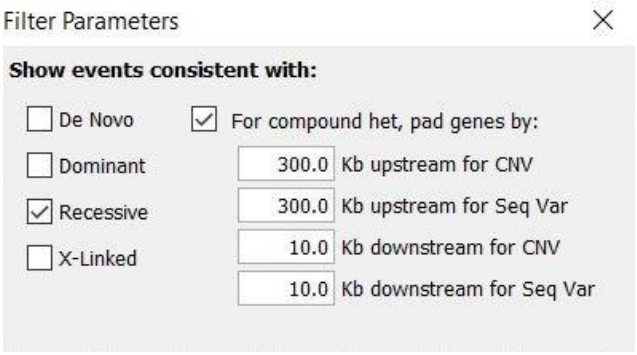


Figure 2706. The inheritance pattern filter.

The **Inheritance Comments** column of the table lists the matching inheritance models. These values are also displayed in the **Variant Details** view using the following abbreviations:

- DN – de novo
- AR – autosomal recessive
- AD – autosomal dominant
- CH – compound het
- XR – X-linked

For duo analysis where the software assumes the missing parent is consistent with the inheritance pattern, the terms **Possibly De Novo** and **Possibly Dominant** are used in the **Inheritance Comments** column.

For sequence variant events, the event matches if both the event type (e.g., SNV, insertion, deletion) and the Alternate Allele match. In contrast, if at the same location, there is an SNV in the proband with Alternate Allele C and the Father has an SNV with Alternate Allele T, this is not a match.

Inheritance Models

***De novo*:** Selecting *De Novo* will only remove events in the Proband that are present in either parent or Unaffected Sibling. If a parent sample is missing, then the event is considered to be the parent having absolutely no variants.

Recessive: For variants to be displayed, the variant must be present in the proband as homozygous and must be heterozygous in parents. For an unaffected sibling (if the sample is present), the variant must either be absent or heterozygous in the sibling to be displayed when this inheritance model is selected.

The filter also shows compound heterozygous events that overlap the same gene - e.g., a loss in father and an SNV in mother on the same gene. In this case the event does not need to be homozygous in the proband, but one variant must be present in each parent. Loss events must overlap the gene, or the padded region as specified in this filter by the Kb upstream and downstream settings for CNV and SeqVar. For example, the settings described above would include any event fully enclosed within the following range: the gene plus CNVs within the 300kb region upstream and 10kb region downstream of the gene, and SeqVar within the 300kb region upstream and 10kb region downstream of the gene. If only one parent is linked to the proband, the filter assumes that the other parent's variant is a pass for this filter. Since this filter affects the status of zygosity and compound events filters, these filters will be disabled (cannot be used) when the inheritance filter is on. When the recessive filter is off, the zygosity and compound events filters will be enabled.

Variants on the sex chromosome are not considered for the recessive filter and will be removed from calculations unless they are on the PAR region, when they will be treated as autosomal.

- **Dominant:** The father or the mother must have the variant present and unaffected siblings must not have the variant at all.
- **X-linked:** The mother must have the variant on the X chromosome.
- **Relationship with compound events and zygosity filters:** These filters are disabled (grayed out) when the Recessive Inheritance Pattern filter is active.
- **Show/hide event types:** The top of the filter panel also houses some buttons to hide/show all CNV, allelic, or sequence variant events. The **Search** bar for the table is also located here.

The buttons in **Figure 277** are toggle switches to hide or show the respective events in the table. When an event type is hidden, the button is displayed with a strikethrough.

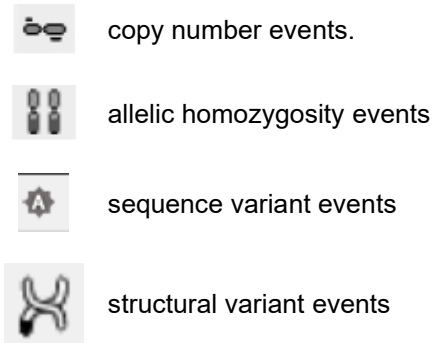


Figure 2717. Toggle switched for respective events.

- **The Search function:** Searching by location or gene symbol using the **Searc**h field limits the rows displayed in the table to regions overlapping the searched location or gene name. To remove this filter, clear the search term in the field and click **Enter.**

The **Filters** tab works in a cascade function. All the filters mentioned above happen one at a time, not all at once. They enter one filter and then anything that makes it through that filter moves on to the next filter. For the copy number and allelic tabs, the events will enter the classified events filter first. For SeqVar, events will enter the ClinVar Classification first. The filters tab shows exactly how many events went into the filter and how many went through. In **Figure 278**, thirty-seven copy number events exited the **Similar Previous Cases** filter and entered the **Event Types** filter. Thirty-seven copy number events exited the **Event Types** filter (showing no filter was applied as the number in equals the number out). However, thirty-seven events entered the Size / No. of Probes filter but only sixteen exited demonstrating that twenty-one events were filtered out by size.

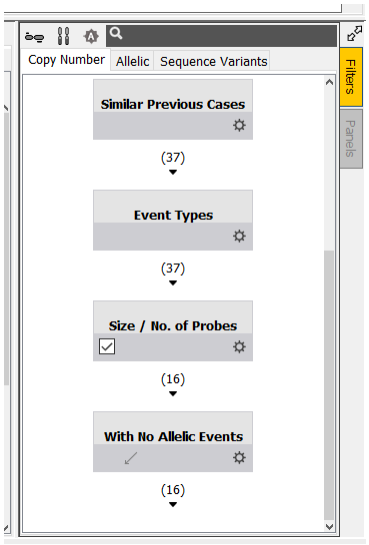


Figure 2728. Cascade function of filters.

Annotated and Unannotated Variant Files

The **SeqVar** tab is used to define processing types for sequence variants based on build and data type. Supported file types are NirvanaJSON, VCF, and VCF Nirvana, shown in **Figure 279**.

- **NirvanaJSON:** Processes input JSON files.
- **VCF:** Processes input VCF files that were annotated through Variant Effect Predictor (VEP). Will also allow loading of unannotated VCF, but the data will remain unannotated.
- **VCF Nirvana:** Annotates the input file via the linked Nirvana annotator and then processes the annotated results.

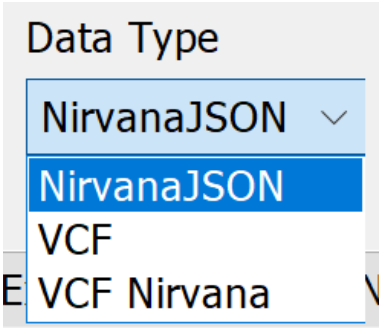


Figure 2739. Supported file types.

Annotated VCF files loaded as Data Type VCF can include annotations from different tools/sources. The following VCF annotations are currently supported in the software.

- Ensembl Variant Effect Predictor
- Nirvana (recommended)

Based on the information available in the **VCF** file header, the software automatically determines the type of annotations present and applies the appropriate parser. If it cannot find a match, then the annotations cannot be loaded.

The NirvanaJSON files are produced via the Nirvana variant annotator from Illumina. More details are available here: https://github.com/Illumina/Nirvana/wiki. In addition, the Nirvana Annotator can be installed separately, and linked by VIA to provide a one-step annotation and interpretation workflow. To use the annotator, select the Data Type VCF Nirvana and load an unannotated VCF file. Once loaded, VIA automatically sends the file to the annotator and then displays the results just like with an annotated VCF or JSON file.

Filtering Events Based on the Filter Column in VCF Files

The VCF file contains a column titled **Filter** which has values that can be used to refine chosen variant parameters. **The VCF Filter Label** parameter in **Settings** allows exclusion or retention of events matching labels found in the FILTER column, depicted by the examples in **Figure 280** below.

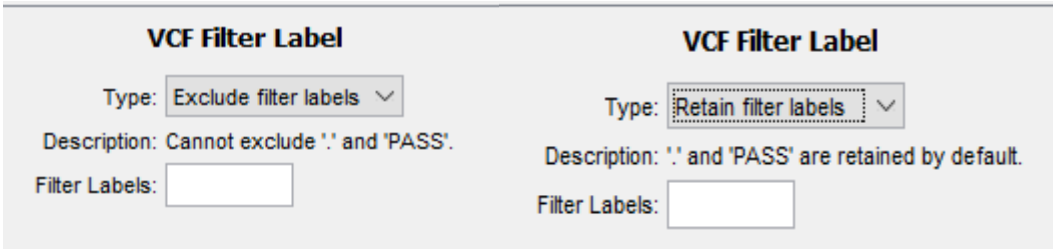


Figure 27480. Examples of Filter Labeling.

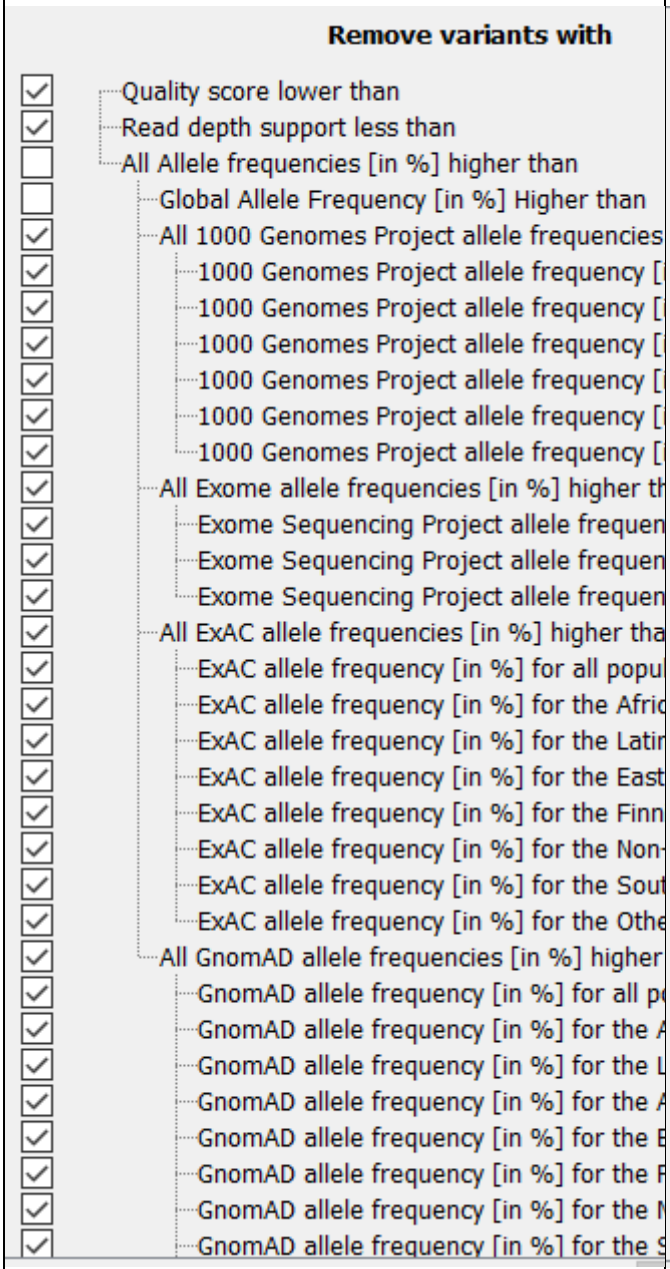
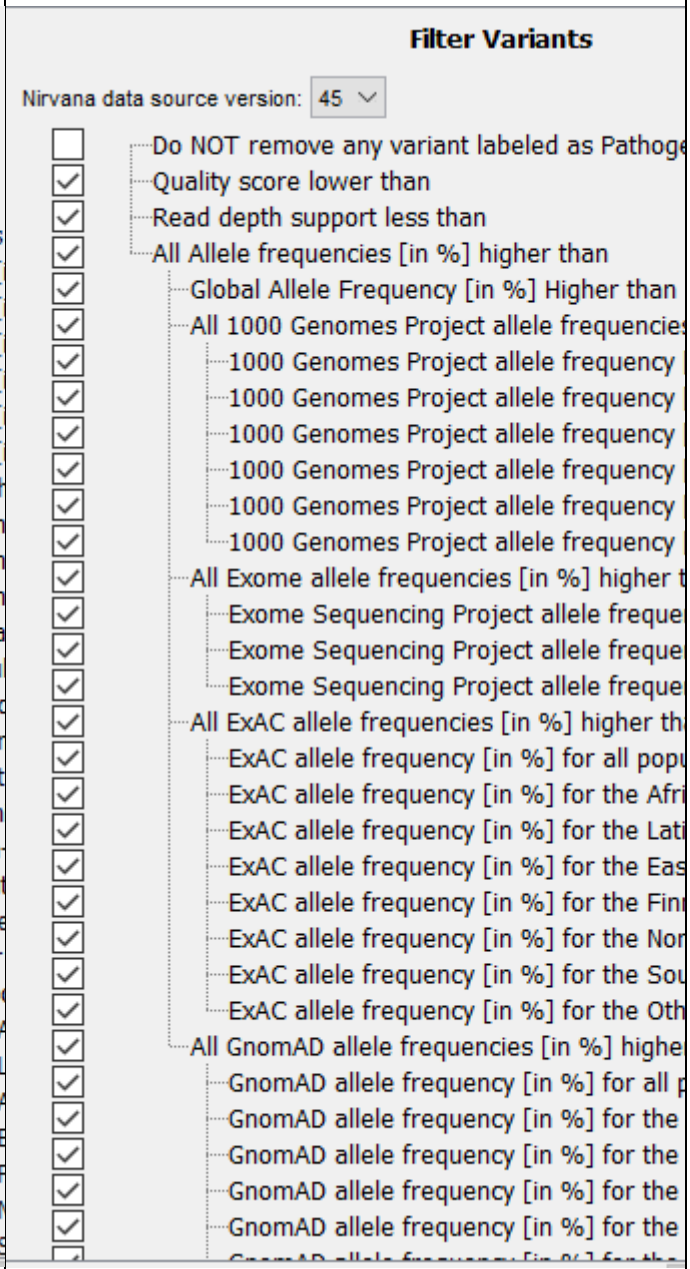
- **Type:** This dropdown determines whether matching variants (as per the values specified in Field Labels) will be excluded or retained. Values: Retain filter labels, Exclude filter labels
- **Description:** This field is just a note stating that variants with the Labels and PASS are automatically retained and cannot be excluded. There is no need to specify these in the filter Label if retaining based on filter values.
- **Filter Labels:** Type in comma separated labels (e.g., OffTarget) used in the FILTER column of VCF files to use for filtering of variants during processing in VIA. Values here are case sensitive and must match exactly the values in the VCF file.

Filtering Events Based on Quality Metrics and Population Frequencies

The processing parameters for NirvanaJSON and VCF Nirvana are the same as both apply to files annotated with the Nirvana annotator. Those for VCF are different, and only results from the VEP annotator are supported by VIA.

Table 18 below helps to delineate the settings for the different file types discussed here and illustrates the VCF Filter labels as well as the Nirvana JSON file labels. VCF Nirvana and NirvanaJSON have additional dropdown fields: Nirvana data source version and a field for ClinVar labeling – when marked, will not remove any variant labeled as Pathogenic during processing and subsequent filtering in the UI.

Table 18. Different File Type Settings.

Settings for VCF:	Settings for VCF Nirvana or NirvanaJSON:
	

NIRVANA DATA SOURCE VERSION

Selection of this version will result in different filter fields displayed:

- Select 45 for files annotated with data source version 45 and lower (linked Nirvana annotator used with VIA version 5.0 and older).
- Select 46 for files annotated with data source version 46 (packaged with the linked Nirvana annotator for VIA version 5.1 and newer).

If during processing the software encounters an incompatibility between the processing type applied to the sample and the data source used to annotate the file, an error will be recorded, and the file will not be processed. An example is given in **Figure 281**.

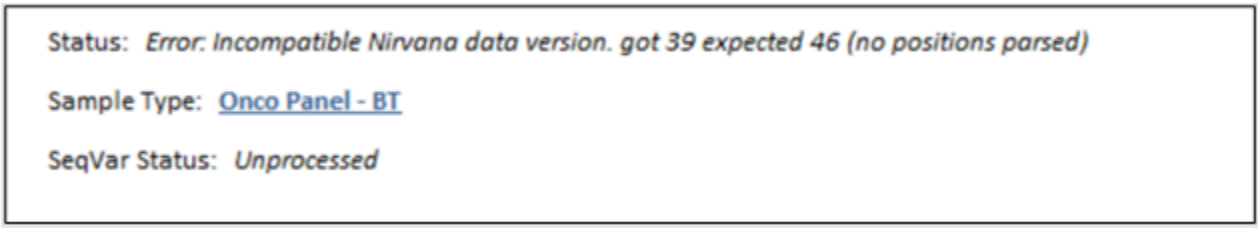


Figure 2751. Example of a processing error.

The error above indicates that the input file was annotated by version 39 of the annotation database using Nirvana, but the processing parameters are specified for version 46 for the uploaded sample. The sample should either be re-annotated with the linked Nirvana Annotator (using the latest databases), or the processing parameters should be changed so that version 45 is selected (older database version) and the correct processing parameters can be applied.

Data source version 46, seen in **Figure 282**, uses newer cohorts and allele frequencies obtained from gnomAD; both the Global and Exome frequencies are included.

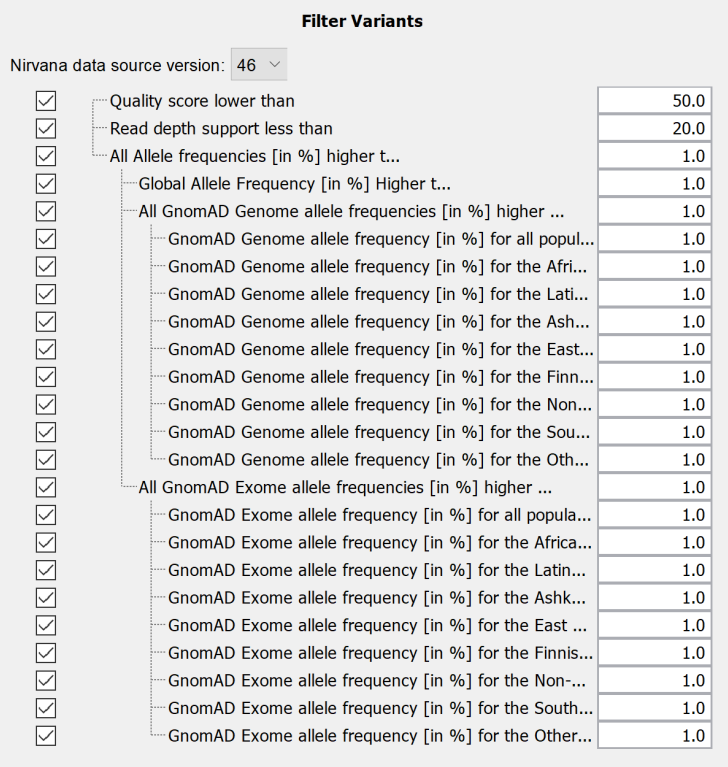


Figure 2762. Data source version 46.

Much of the data from ExAC is incorporated within the new gnomAD frequencies (data source version 46). For detailed information refer to the gnomAD website (About and FAQs).

Version 45 contains older data sources and cohorts and should be applied to legacy samples processed through an older Nirvana annotator. Nirvana JSON output files contain data source version numbers, and this is the field parsed by VIA to determine compatibility.

Reviewing OGM Data

OGM sample review

HOME PAGE

Once the ogm sample has been processed, the sample can be queried on the **Home** page. Details of the sample will appear on the **Home** page, including QC metrics.

A user may launch the Access Circos plot from the **Home** page by clicking on the **Circos Plot** icon next to sample information (**Figure 283**). Clicking on the icon will launch **Access** on the browser to display the Circos plot.

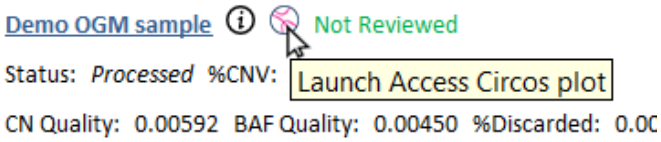


Figure 2773. Access Circos Plot.

CIRCOS PLOT

When opening an OGM sample for review, the first tab in the top panel is the Circos plot. In this plot, the chromosomes are displayed in a circular pattern. Translocations are displayed as colored arcs on the Circos plot with each translocation having a different (ascending rainbow) color. When the mouse is placed over one of the translocation lines, a tool tip appears showing the ISCN representation on the first line and the genes on each end on the translocation on the second line, seen in **Figure 284**.

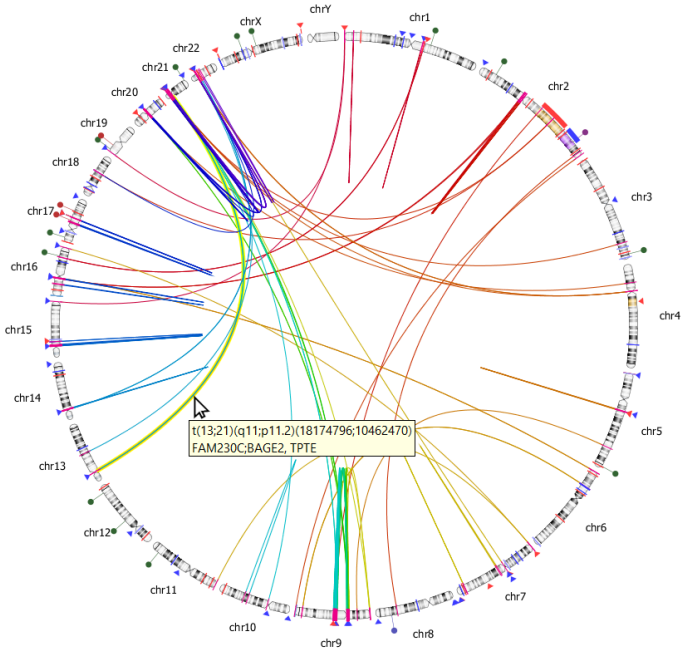


Figure 2784. Tool tip.

The genes on the upstream or lower chromosome number end of the translocation are listed first, followed by genes in the downstream or higher chromosome numbers. A semicolon (;) separates between the 2 ends of the translocation and a comma (,) is used to list multiple genes on each end. If there are no genes at either end, then it will remain blank.

As shown in the example image of a circle plot zoomed into chr 2 in **Figure 285**, copy number gains and losses are visible on the outer side of the chromosome circle plot as blue and red arrows or bars, respectively. Additionally, allelic events, such as AOH, and UPD events are observed as shaded regions inside the chromosomes. AOH is shaded yellow, allelic imbalance is purple, and UPD events have mixed shading matching the **Tracks Overview** tab. Sequence variant events processed with the sample will appear as lollipops on the outer side of the chromosome.

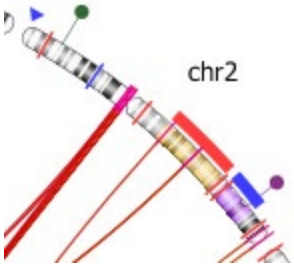


Figure 2795. Circos plot zoomed.

Applying filters to the events will update the Circos plot such that the events in the Circos plot matches the events in the event table. Clicking on the chromosome image in the Circos plot will switch the view to the tracks panel of the sample. Adjusting the window of the circle plot will resize the image.

STRUCTURAL VARIANT TRACK

In the **Tracks** tab of the **Table & Tracks Preferences** window, there are several structural variant specific tracks that can be turned on for display. SV Events, OGM Molecules, and OGM Coverage are tracks specific to structural variants. These tracks can be enabled by checking the box under **Show** in **Track Preferences**.

SV EVENTS

The SV Events track displays structural events in the **Tracks** panel. The events in this track originate from the SV Data Type. The event types that are displayed in the **SV Events** track include deletion, insertion, interchromosomal translocation, intrachromosomal fusion, inverted duplication, tandem duplication, unplaced duplication, inversion, partial inversion, and paired inversion.

An event in the **SV Event** track is displayed with a horizontal bar for each side of the break end region. A vertical black bar in the middle of the horizontal bar indicates the midpoint of the break end region. An example SV event is shown in **Figure 286**.

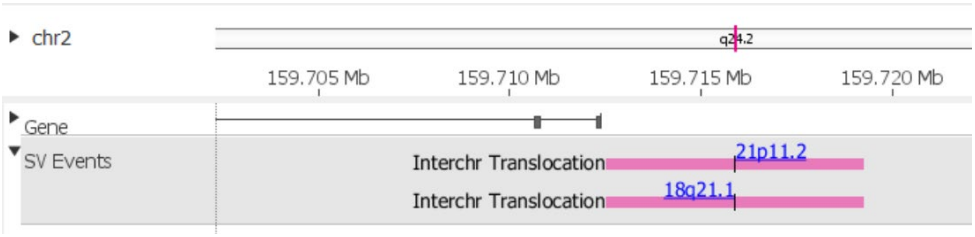


Figure 2806. SV Event track.

The SV event type appears to the left of the event. For paired break ends (e.g., Interchr Translocation) the cytoband label will appear above the paired when the track is zoomed in. The cytoband label can be clicked to switch the track view to the paired break end.

OGM MOLECULES

The OGM Molecules track will display the molecules from the ogm.bam file in samples processed with CNV from either the OGM BAM Multiscale or OGM BAM Self-Reference data types. Users will need to zoom into a region for the OGM Molecules track to appear. The OGM molecules are colored according to the key indicated in **Figure**

287. The OGM Molecules coloring key can be visible in a pop out window when clicking on the **?** icon under the **OGM Molecules** track label.

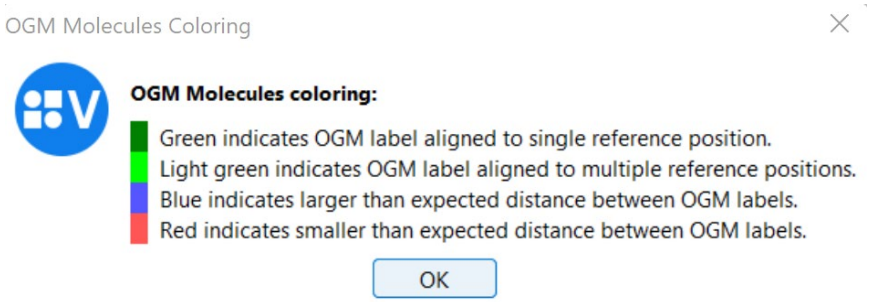


Figure 2817. OGM Molecules track color key.

When the track is zoomed in to the nucleotide level, information on the individual molecule will appear in the tool tip by placing the mouse over each molecule. The Read Name, CIGAR, and Mapping Quality appear in the tool tip.

OGM COVERAGE

The OGM Coverage track displays the coverage depth of the OGM molecules. The units on the left side of the track indicate the number of OGM molecules and the black bars provide the depth of coverage. The yellow line across the track indicates the median depth of the viewable region, as shown in **Figure 288**.

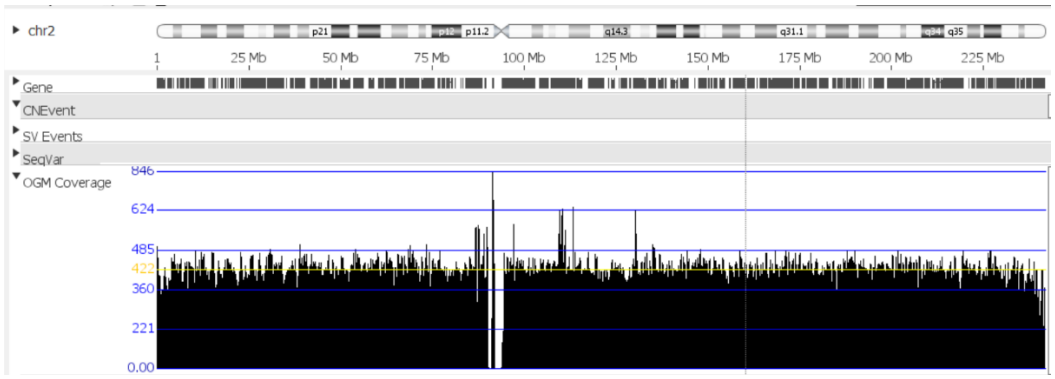


Figure 2828. OGM Coverage track.

SV EVENTS FILTERS

The **Structural Variant Events** filters are accessible under the **Filters** tab to the right of the table. The entire **SV Events** filter pipeline is shown in **Figure 289**.

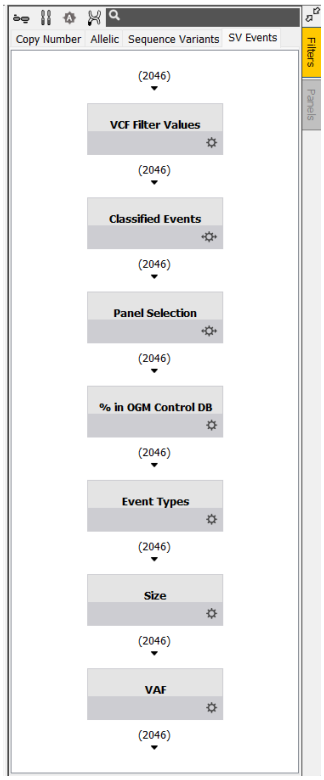


Figure 2839. SV Events filter pipeline. For quick on/off filtering of all SV Events, click the icon on the top menu of the filter.

VCF Filter Values: The ability to filter SV events annotated as PASS, Low Confidence, Masked and Poor Molecule Support by Solve are available in the **VCF Filter Values** filter. For additional description of these values, please see *Bionano Solve Theory of Operation: Structural Variant Calling* (P/N 30110).

Classified Events: Applied to filter according to the event classification. This filter is the same for all CN filter pipelines. A detailed description of the **Classified Events** filter can be found in the *Filtering of CNV, Allelic Events, and Sequence Variant Data* section of the *VIA User Guide*.

Panel Selection: Filters according to a gene or region panel. A detailed description of the **Panel Selection** filter can be found in the *Filtering of CNV, Allelic Events, and Sequence Variant Data* section of the *VIA User Guide*.

% in OGM Control DB: Filtering SV events by frequency in the OGM Control can be done in the **% in OGM Control DB** filter, seen in **Figure 290**. For details on the OGM Control DB see *Bionano Solve Theory of Operation: Variant Annotation Pipeline* (P/N 30190).

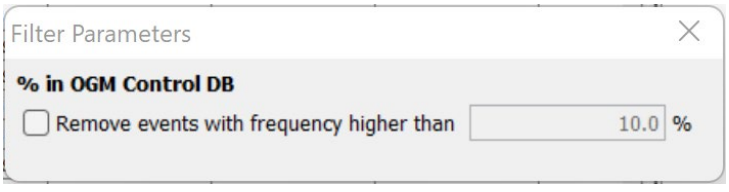


Figure 284. Filter Parameters % in OGM Control DB.

Event Types: Filtering by SV event types (insertions, deletions, tandem and split duplications, and inverted duplications), is available in in the **Event Types** filter, shown in **Figure 291**.

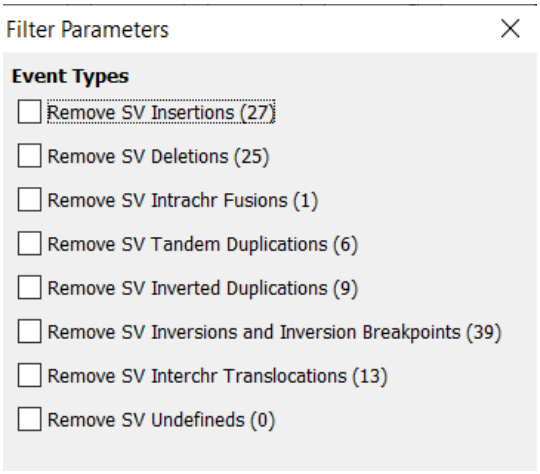


Figure 285. Filter Parameters Event Types.

Size: The ability to filter by the size of SV event insertions, deletions, tandem and split duplications, and inverted duplications is available in in the **Size** filter, shown in **Figure 292**.

NOTE: The values field for the size filter is in kb.

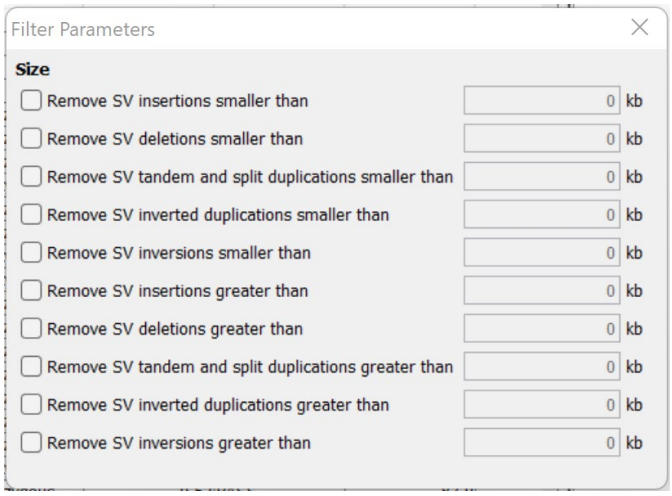


Figure 286. Filter Parameters Size.

VAF: Enables the ability to remove SV events with a specified variant allele frequency. Events in this filter can be removed with a VAF less than a specified value and/or SV events with a VAF greater than % value as shown in **Figure 293**.

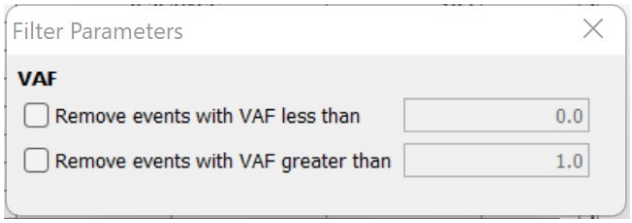


Figure 287. Filter Parameters VAF.

TABLE PREFERENCES

In the **Table** and **Track Preferences** window, several columns are available to be displayed in the table columns. These columns are found in the Structural Variant Events as shown in **Figure 294** and include Fusion Junction 1, Fusion Junction 2, SV Quality, Molecule Count, VCF filter values, % in OGM Control DB, SV VAF, and Zygosity. The events in the table may be sorted by clicking on the column header in the table.

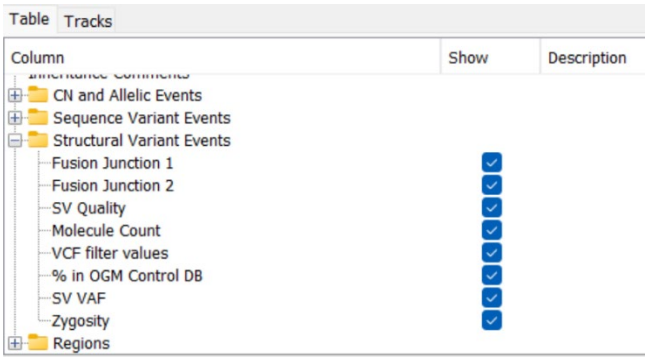


Figure 288. Select/Show Columns.

Fusion Junction 1 and Fusion Junction 2: A fusion junction is the region where two break ends are joined. The break end region(s) are listed under the Fusion Junction columns. For SV Events with only one fusion junction (e.g., Deletion), only Fusion Junction 1 will be populated. For SV Events with 2 fusion junctions, (e.g., Insertion), the second fusion junction is listed in Fusion Junction 2. The direction of the break end region is indicated by the arrow next to the specified region. For fusion junctions with more than one break end, the break end regions are separated by a semi-colon (;).

SV Quality: If an event has an SV Quality score, it will be listed in this column. For more information on SV Quality please see *Bionano Solve Theory of Operation: Structural Variant Calling* (P/N 30110).

Molecule Count: Molecule Count column indicates the number of molecules that support the event.

VCF filter values: SV events annotated as PASS, Low Confidence, Masked and Poor Molecule Support by Solve are available in the VCF filter values column.

% in OGM Control DB: Displays the frequency in percent of the SV event in the OGM Control DB.

SV VAF: Indicates the variant allele frequency for the SV event.

Zygosity: The zygosity of SV events are listed as homozygous, heterozygous, or hemizygous. It is currently assigned to only insertions, deletions, translocations, and inversions.

SV EVENT VARIANT DETAILS

When an SV Event is selected from the table or track, the user may switch the view of the top panel to the **Variant Details** by clicking on the tab. The layout and information displayed is shown in **Figure 295**.

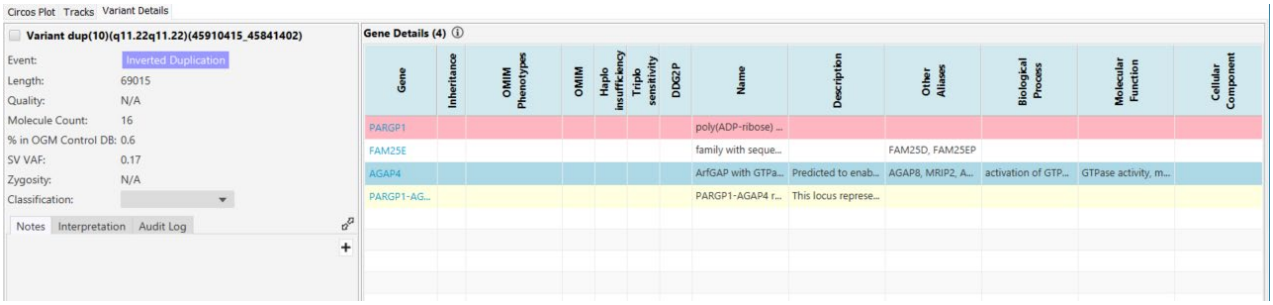


Figure 289. Event Variant details.

The upper left side of the Variant Details contains information about the SV Event. This includes the ISCN nomenclature of the event, event type, length, quality, molecule count, % in OGM Control DB, SV VAF, zygosity and event classification. Below the event details are tabs for **Notes**, **Variant Interpretation**, and **Audit Log**. This section can be detached into another window by clicking on the double arrows .

The right panel contains details of the genes affected by the SV Event. Genes highlighted as blue are genes in the break end zone. Genes highlighted in red are genes in break end two. Genes highlighted in yellow span both break ends. Genes not highlighted are genes within the 25 kb region of the SV Event.

System Administration

The Administrator establishes user accounts and many other features during installation and continued software maintenance.

The Administrator account has an additional **Admin** tab in the window from which users, reports, regions, platforms, processing types, and sample types can be managed. The Administrator login should only be used to set up the system and enforce procedures. It should not be used as a regular user account. For actual processing and review of samples, make sure to log in with a standard user account.

NOTE: The **BAM References** tab is only available for licenses that include NGS.

Login: Upon launching the system, the **Login** screen will appear. There is an option for the Admin account to either log in as an admin or as a regular user. Make sure the appropriate selection is made (checkbox marked = Admin; this box is unchecked by default). The default credentials for the Administrator account are:

Username: admin
Password: admin7

Make sure to change the password after logging in for the first time being certain to record the new password. The password can be changed using the icon on the top right of the **Home** screen, as seen in **Figure 296,** or by using **Modify User** from the **Users** tab.

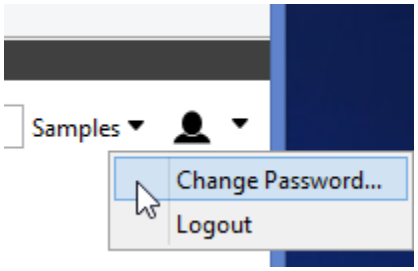


Figure 290. Change password.

The **Repository** button on the top left shows settings for the server where data is stored, seen in **Figure 297**. This is set up during installation but if changes need to be made, the Host and Port can be edited here. A more secure connection can be made by checking the **Use https** box, which will encrypt and decrypt transferred data. Please review the installation guide for details.

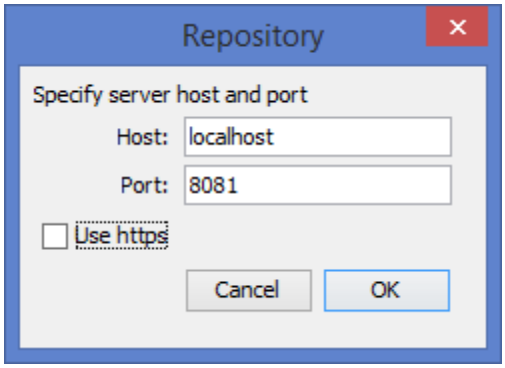
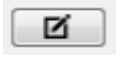


Figure 291. Repository button.

User Accounts: User accounts are managed via the **Users** tab.

NOTE: The user admin does not have any details filled out initially. Please modify information for the admin account upon first login.

Creating a user account: VIA user accounts are created using the following tools located at the bottom of the window:

 Add user  Modify user.

Click on the **Add user** button on the bottom left of the window. Enter a username, password, real name, email address and status for the account. Next assign user privileges by checking the appropriate boxes. User privileges include:

- **Ability to perform admin operations** allows the user to perform all functions of the Admin account ***excluding*** the ability to update software and annotations.

- **Ability to edit sample attributes** allows the user to load, associate and edit sample attributes (factor values)
- **Ability to save view preferences for a sample** allows the user to save view preferences for the single sample specific view, **This Sample View**.
- **Ability to edit genomic events** allows the user to edit events by modifying boundaries, adding, or deleting events.
- **Ability to edit and delete notes** allows user to make notations about an event in the Notes column of the results table and delete notes, as necessary. This permission is only available if Ability to edit and delete notes is checked.
- **Ability to process samples**: allows user to load and process samples.
- **Ability to delete samples**: allows the user to delete samples from the database.
- **Ability to lock samples**: allows the user to lock a sample after a report is generated so that no other changes can be made to the sample.
- **Ability to submit to the KB**: allows the user to submit entries for inclusion in the KB; samples enter the Pending state upon submission.
- **Ability to approve KB submissions**: allows the user to edit and approve Pending KB entries.
- **Ability to delete KB entries**: allows user to delete entries in the KB.

Finally, clicking on **Add User** will add the new user to the system. This user will now appear in the users list in the **User** tab.

NOTE: multiple users can have Admin privileges. The Admin privilege is granted to the first Admin user to log in, preventing any other Admin user from logging in with Admin privileges while one Admin account (with Admin privileges enabled) is in use. In such cases, the other Admin users may login without Admin privileges by unchecking the box in the login window.

Inactivating a user account: Once a user account is created within the system, it cannot be deleted and can only be inactivated. This ensures that any sample/event notes or modifications made by the user will remain in the system and will not be lost. If a user should no longer be allowed to use VIA, the account status can be modified to Inactive, and that user will no longer be able to log in.

To inactivate a user, click on the **Admin** tab and then the **Users** tab. Click on a user row in the table to highlight it and then click on the **Modify User** button. The **Modify User** window will pop up containing user details and privileges. Select **Inactive** for the **Status** field and click the **Modify User** button to save the updated user settings.

To modify a user, click on the **Admin** tab and then the **Users** tab. Click on a user row in the table to highlight it and then click on the **Modify User** button . The **Modify User** window will pop up containing user details and privileges. Make the appropriate changes to the user profile and click the **Modify User** button to save the updated user settings.

Changing a user's password: If a user's password requires changing, this can be accomplished in the **Modify User** window for the individual. Check the **Change Password** checkbox. The **Password** fields will now become active. Enter the new password and click the **Modify User** button to save the new password.

Regions: The Administrator maintains all the annotation files and the reference data collection that is to be used within the VIA system. In the **Regions** tab, the Administrator can upload tracks and the annotation/regions files

can be grouped together in folders. For example, there can be different regions/tracks for Cancer, Constitutional, Benign, and Pathogenic and these categories can be organized by creating such folders to house the relevant files in each. Each genomic build has its own set of annotation files.

The **BioDiscovery Provided Regions** folder houses region files that are regularly updated by Bionano. Any user added tracks will be housed in the **Custom Regions** folder. See **Figure 298** for an illustration.

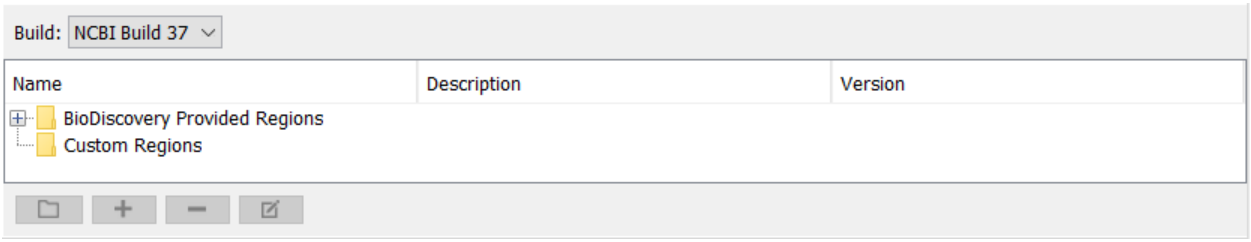


Figure 292. Build folders.

Annotation file format: Annotation/regions files must be .BED or .TXT files to be loaded into the system.

BED files: For this file format, please refer to https://genome.ucsc.edu/FAQ/FAQformat.html. The file can contain a header line and at minimum must contain columns and values for **chromosome**, **start position**, and **end position**. An example of a BED file is shown in **Figure 299**.

```
# Sequence  Min Max Name
track name="DECIPHER_Syndromes_hg19" description="DECIPHER_Syndromes_hg19 Jul 24, 2
chr1    1    4837854 1p36_microdeletion_Syndrome
chr1    145413190    147465755    1q21.1_Thrombocytopenia-Absent_Radius_(TAR)_Syndrom
chr1    146512930    147737500    1q21.1_recurrent_microduplication_(possible_suscept
chr1    146512930    147737500    1q21.1_recurrent_microdeletion_(susceptility_locus_
chr2    57741796     61738334     2p15-16.1_microdeletion_Syndrome
chr2    196925089    205206940    2q33.1_deletion_Syndrome
chr2    239969863    243199373    2q37_monosomy
chr3    195672229    197497869    3q29_microduplication_Syndrome
chr3    195672229    197497869    3q29_microdeletion_Syndrome
chr4    1    2230958 Wolf-Hirschhorn_Syndrome
chr5    10001    11723854     Cri_du_Chat_Syndrome_(5p_deletion)
chr5    112101596    112221377    Familial_Adenomatous_Polyposis
chr5    126063045    126204952    Adult-onset_autosomal_dominant_leukodystrophy_(ADLD
chr5    175130402    177456545    Sotos_Syndrome
chr6    391760  1312675 6p25deletion_Syndrome
chr7    72332743     74616901     7q11.23)duplication_Syndrome
chr7    72332743     74616901     Williams-Beuren_Syndrome_(WBS)
chr7    95533860     96779486     Split_hand/foot_malformation_1_(SHFM1)
chr8    8119295 11765719     8p23.1_dup/deletion_Syndrome
```

Figure 293. An example of a BED file.

Another example, shown in **Figure 300**, contains hyperlinks to the Decipher website. When displayed in VIA, and by moving the mouse over regions in this track, the pointer will turn into a hand indicating that it is a hyperlinked region. Clicking on the region will open the relevant page on the Decipher website.

```
track name=Decipher_Syndromes type=bedDetail description="Decipher_Syndromes 2015-06-11 track" db=hg19 visibility=3
url="https://decipher.sanger.ac.uk/syndrome/$$"
chr4    1569197 2110236 Wolf-Hirschhorn Syndrome    0    .    1569197 2110236 0,0,0              1    <html><head></head><body><hr><a
href="https://decipher.sanger.ac.uk/syndrome/1" target="_blank">Wolf-Hirschhorn Syndrome
details</a></hr><br>0.54</br><br>1</br><br>1</br></body></html>
chr5    10001   12533304    Cri du Chat Syndrome (5p deletion)  0    .    10001   12533304    0,0,0              2    <html><head></he
href="https://decipher.sanger.ac.uk/syndrome/2" target="_blank">Cri du Chat Syndrome (5p deletion)
details</a></hr><br>12.52</br><br>1</br><br>2</br></body></html>
chr7    72744455    74142672    Williams-Beuren Syndrome (WBS)  0    .    72744455    74142672    0,0,0              3
<html><head></head><body><hr><a href="https://decipher.sanger.ac.uk/syndrome/3" target="_blank">Williams-Beuren Syndrome (WBS)
details</a></hr><br>1.4</br><br>1</br><br>3</br></body></html>
chr15   22749354    28438266    Angelman syndrome (Type 1)  0    .    22749354    28438266    0,0,0              4    <html><head></he
href="https://decipher.sanger.ac.uk/syndrome/4" target="_blank">Angelman syndrome (Type 1)
details</a></hr><br>5.69</br><br>1</br><br>4</br></body></html>
chr16   3775055 3930121 Rubinstein-Taybi Syndrome http://www.ncbi.nlm.nih.gov/books/NBK1526/    0    .    3775055 3930121 0,0,0
<html><head></head><body><hr><a href="https://decipher.sanger.ac.uk/syndrome/7" target="_blank">Rubinstein-Taybi Syndrome
http://www.ncbi.nlm.nih.gov/books/NBK1526/ details</a></hr><br>0.16</br><br>1</br><br>7</br></body></html>
chr17   16773072    20222149    Smith-Magenis Syndrome  0    .    16773072    20222149    0,0,0              8    <html><head></head><
href="https://decipher.sanger.ac.uk/syndrome/8" target="_blank">Smith-Magenis Syndrome
details</a></hr><br>3.45</br><br>1</br><br>8</br></body></html>
```

Figure 294. Decipher website links.

Another Decipher BED file opened in MS Excel to show as columns (total of 14) is seen in **Figure 301**. To enable html links, they must be placed in column 14 and the BED file must contain fourteen columns; the first four columns (A-D) are used as well as column I (color specs in RGB; in the example below 0,0,0 specifies the color black) and column N (html hyperlink).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	track name=Decipher_Syndromes_Duplications type=bedDetail description="Decipher_Syndromes Duplications 2017-09-29 track" db=hg19 visibility=3 url="https://decipher.sanger.ac.uk/syndrome/$$"																			
2	chr17	16869758	20318836	Potocki-Lu		0 .	16869758	20318836	0,0,0				19	<html><head></head><body><hr><a href="https://decipher.sanger.ac.uk/sy						
3	chr17	14194598	15567589	Charcot-M		0 .	14194598	15567589	0,0,0				29	<html><head></head><body><hr><a href="https://decipher.sanger.ac.uk/sy						
4	chrX	1.04E+08	1.04E+08	Pelizaeus-M		0 .	103776510	103792618	0,0,0				38	<html><head></head><body><hr><a href="https://decipher.sanger.ac.uk/sy						
5	chr7	73330452	74728334	7q11.23 du		0 .	73330452	74728334	0,0,0				43	<html><head></head><body><hr><a href="https://decipher.sanger.ac.uk/sy						
6	chrX	1.54E+08	1.54E+08	Xq28 (MEC		0 .	154021812	154097731	0,0,0				45	<html><head></head><body><hr><a href="https://decipher.sanger.ac.uk/sy						

Figure 295. BED file opened in Excel.

The BED format is quite flexible. The column headers can be in any row if it exists in the file, the file can have a header line, and it can also have comments lines (those starting with a hash tag). For example, the following formats are all acceptable:

- A header line and the column headers present after the first line of data:

```
track name="DECIPHER_Syndromes_hg19" description="DECIPHER_Syndromes_hg19 Jul 24, 2013 8:1
chr1    1    4837854 1p36_microdeletion_Syndrome
Sequence    Min Max Name
chr1    145413190    147465755    1q21.1_Thrombocytopenia-Absent_Radius_(TAR)_Syndrome
chr1    146512930    147737500    1q21.1_recurrent_microduplication_(possible_susceptiblity_
chr1    146512930    147737500    1q21.1_recurrent_microdeletion_(susceptility_locus_for_neu
chr2    57741796     61738334     2p15-16.1_microdeletion_Syndrome
chr2    196925089    205206940    2q33.1_deletion_Syndrome
```

- No header line with column headers as the first line in file:

```
Sequence    Min Max Name
chr1    145413190    147465755    1q21.1_Thrombocytopenia-Absent_Radius_(TAR)_Syndrom
chr1    146512930    147737500    1q21.1_recurrent_microduplication_(possible_suscept
chr1    146512930    147737500    1q21.1_recurrent_microdeletion_(susceptility_locus_
chr2    57741796     61738334     2p15-16.1_microdeletion_Syndrome
chr2    196925089    205206940    2q33.1_deletion_Syndrome
```

- No header line and no column headers:

```
chr1    145413190   147465755   1q21.1_Thrombocytopenia-Absent_Radius_(TAR)_Syndrome
chr1    146512930   147737500   1q21.1_recurrent_microduplication_(possible_susceptibl
chr1    146512930   147737500   1q21.1_recurrent_microdeletion_(susceptility_locus_for
chr2    57741796    61738334    2p15-16.1_microdeletion_Syndrome
chr2    196925089   205206940   2q33.1_deletion_Syndrome
```

TXT Files: For annotation files in txt format, column headers are required and at minimum must have the following columns: **Chromosome, Start,** and **End.**

The TXT format is not as flexible as the BED file format. The header line in a TXT file must be the first line of the file and subsequent lines must be the values. Additional columns may be present in the file but will be ignored by VIA. An example annotation text file (opened in MS Excel) containing Chromosome, Start, and End columns is shown in **Figure 302** and an example of an additional column is shown in **Figure 303**.

	A	B	C
1	Chromosome	Start	End
2	chr1	2326240	2354009
3	chr1	2975743	3365184
4	chr1	214521010	214734641
5	chr1	6235079	6269678
6	chr1	1839028	1860739
7	chr1	3559128	3660466
8	chr1	3537330	3576670
9	chr2	44056102	44115604
10	chr2	56401257	56623308
11	chr2	62122802	62373204
12	chr2	38284745	38313322
13	chr2	241055979	241085763
14	chr2	63267964	63294313
15	chr2	233402778	233425225
16	chr2	71117719	71170575
17	chr2	43439540	43463744
18	chr3	125812407	125909484
19	chr3	193843933	193866395

Figure 296. A TXT files.

	A	B	C	D
1	Chromosome	Start	End	Gene
2	chr1	40213902	40264532	BMP8B Paternal Predicted
3	chr1	68501644	68526459	DIRAS3 Paternal Imprinted
4	chr1	1260657	1294491	DVL1 Maternal Predicted
5	chr1	24161566	24204820	FUCA1 Paternal Predicted
6	chr1	92930317	92962432	GFI1 Paternal Predicted
7	chr1	226702430	226722881	HIST3H2BB Maternal Predicted
8	chr1	161484035	161506686	HSPA6 Maternal Predicted
9	chr1	108037749	108058249	NDUFA4P1 Paternal Predicted
10	chr1	228385860	228576574	OBSCN Paternal Predicted
11	chr1	247994229	248015197	OR11L1 Paternal Predicted
12	chr1	2326240	2354009	PEX10 Maternal Predicted
13	chr1	2975743	3365184	PRDM16 Paternal Predicted
14	chr1	214521010	214734641	PTPN14 Maternal Predicted
15	chr1	6235079	6269678	RPL22 Paternal Predicted
16	chr1	1839028	1860739	TMEM52 Paternal Predicted
17	chr1	3559128	3660466	TP73 Maternal Imprinted
18	chr1	3537330	3576670	WDR8 Maternal Predicted
19	chr2	44056102	44115604	ABCG8 Maternal Predicted
20	chr2	56401257	56623308	CCDC85A Paternal Predicted

Figure 297. A TXT file containing an additional column which is ignored by VIA.

Tools for Managing Annotation Files: Tools on the bottom left of the window allow creation of folders and addition of region files. Creating folders/subfolders: For better organization of annotations files (tracks), folders and subfolders can be created for each genomic build within the Custom Regions folder. Users cannot add/delete folders/files in the BioDiscovery Provided Regions folder.

Make sure the correct build is selected in the **Build** dropdown field, click the **Custom Regions** folder, and click the **Add Subfolder To** button. Provide a name in the popup window and a new folder will be created.

Adding a new Region file: Highlight the folder in which the new region file should be placed and click on the **Add Regions** button on the bottom left. Select the file and provide additional information for the data being loaded, as seen in **Figure 304**.

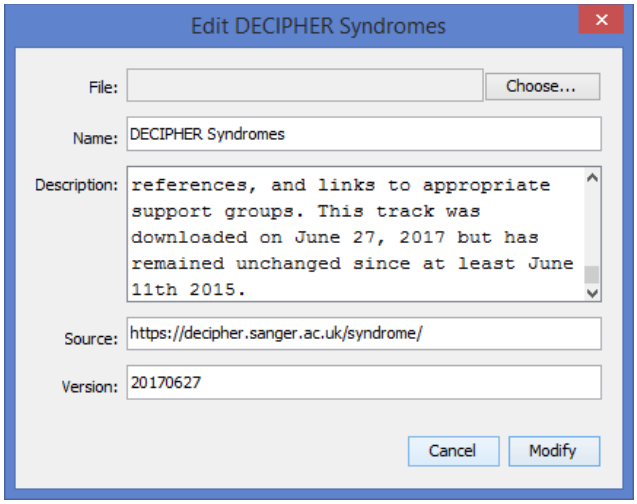


Figure 298. Creating Region files.

After the file is loaded and the track is created, it will be visible in the **Regions** tab within the selected folder, as shown in **Figure 305**.

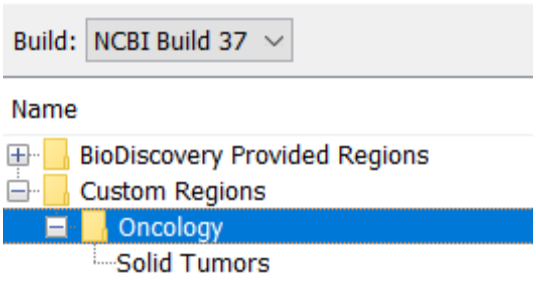


Figure 299. Build is displayed.

Modifying the track name: The region name and other information can be changed by highlighting the track name and clicking on the **Edit** button at the bottom of the window. This will bring up the **Edit** window where values can be changed. Once changes are made, click on **Modify** to save the new values.

Deleting folders/tracks: Click on a folder or track name to select it – the row should be highlighted in dark gray. To select multiple folders/tracks, hold down the **CTRL** button while clicking on folders/tracks. Then click on the **Delete selected item(s)** button. To delete multiple tracks at a time, all selected tracks must be in the same folder. Note that only empty folders can be deleted. To delete a folder containing tracks, first delete all tracks then the folder itself.

NOTE: Items in the **BioDiscovery Provided Regions** folder cannot be deleted.

Platforms: Within the **Platforms** tab, the Administrator creates protocols for processing samples. A set of processing parameters pertain to a specific genome build, data type, manufacturer (if applicable) and assay name (if applicable). A processing type name is given to each settings profile that is defined.

There are two tabs here: **CNV** and **SeqVar**, one for each type of data modality (copy number variation or sequence variation, respectively). Within each VIA installation, each provided Assay comes with one or more example processing types. These processing type names are preceded by Example.

NOTE: the example processing types that come with the installation cannot be edited (all fields will be grayed out) and cannot be applied to a sample type for processing. First make a copy of that processing type, rename it, and apply the processing type to the sample type.

CNV: Different settings parameters are available based on the data type and associated factors. In **Figure 306**, the processing type for the Illumina CytoSNP850k arrays includes settings applicable to SNP arrays (e.g., Homozygous Frequency Threshold, Minimum LOH Region (KB)) whereas for the Agilent 4x180k arrays, seen in **Figure 307**, such settings are not there because it is not an SNP array.

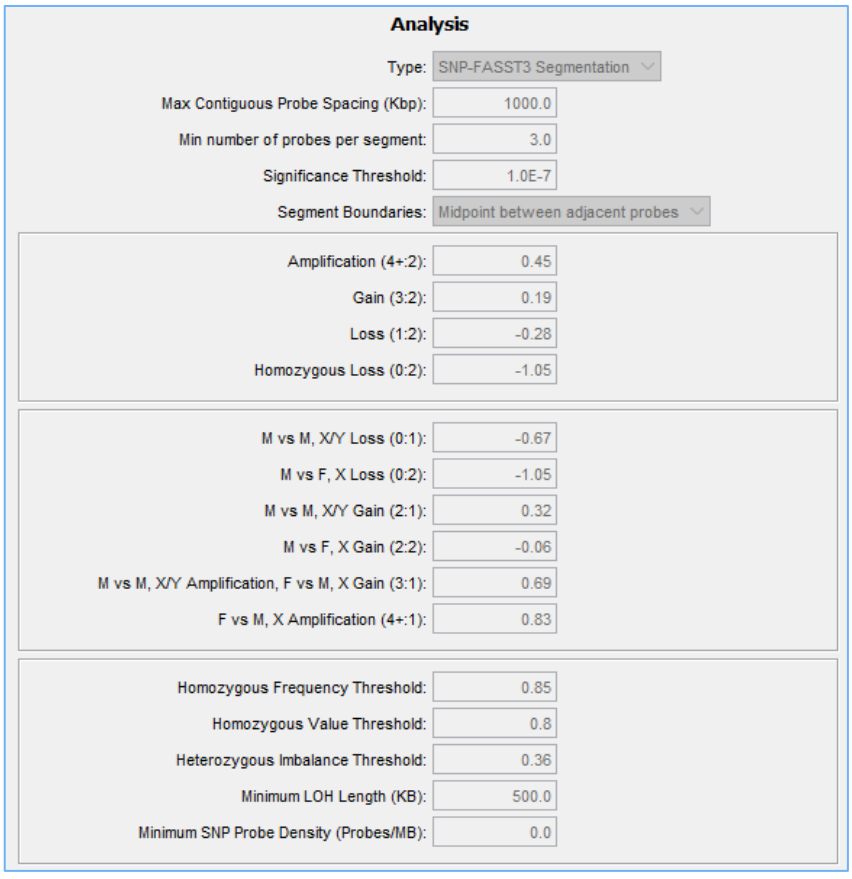


Figure 300. Illumina Infinium CytoSNP 850K arrays.

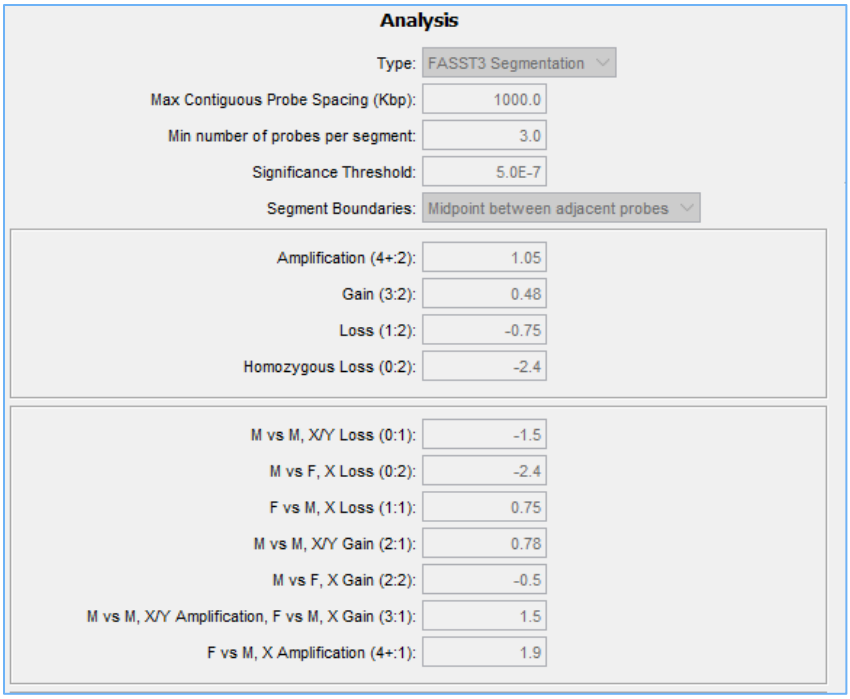


Figure 301. Agilent SurePrint CGH 4x 180k arrays.

Table 19 below describes each of the configurable processing settings for event segmentation.

Table 19. Description of event segmentation processing settings.

Processing Setting	Description
Significance Threshold	Sets statistical threshold for producing a state change
Max Contiguous Probe Spacing (Kbp)	Gap size between probes/datapoints before the segment is ended
Min number of probes per segment	Number of probes to produce a segment
Amplification	Log2 value for multiple copy Gain
Gain	Log2 value for a single copy Gain
Loss	Log2 value for heterozygous Loss
Homozygous Loss	Log2 value for nullizygous Loss
M vs M, X/Y Loss (0:1)	Log2 value for hemizygous Loss of sex chromosomes for Males when Control gender is Male
M vs F, X Loss (0:2)	Log2 value for nullizygous Loss of chromosomes X for Males when Control gender is Female
F vs M, X Loss (1:1)	Log2 value for Loss of chromosome X for Females when Control gender is Male

M vs M, X/Y Gain (2:1)	Log2 value for single copy Gain of sex chromosomes for Males when Control gender is Male
M vs F, X Gain (2:2)	Log2 value for single copy Gain of chromosome X for Males when Control gender is Female
M vs M, X/Y Amplification, F vs M, X Gain (3:1)	Log2 value for multiple copy Gain of sex chromosomes for Males or single copy Gain of chromosome X for Females when control gender is Male
F vs M, X Amplification (4:1)	Log2 value for multiple copy Gain of chromosome X for Females when control gender is Male
Homozygous Frequency Threshold	Percentage of homozygous SNPs needed to generate an AOH/LOH event
Homozygous Value Threshold	BAF value for genotyping a SNP as homozygous
Heterozygous Imbalance Threshold	BAF value for genotyping a SNP as heterozygous
Minimum LOH Length (KB)	Minimum segment size to generate an AOH/LOH event
Minimum SNP Probe Density (Probes/MB)	Minimum number of SNP datapoints needed to generate an AOH/LOH event

CNV from NGS – BAM MultiScale: The processing type for calling copy number from NGS samples (BAM to CNV analysis) using a reference file is found under the Data Type BAM Multiscale found at the bottom of the **Data Type** dropdown, shown in **Figure 308**. Note this UI does not have the **Manufacturer** and **Assay Type** fields as they are not applicable.

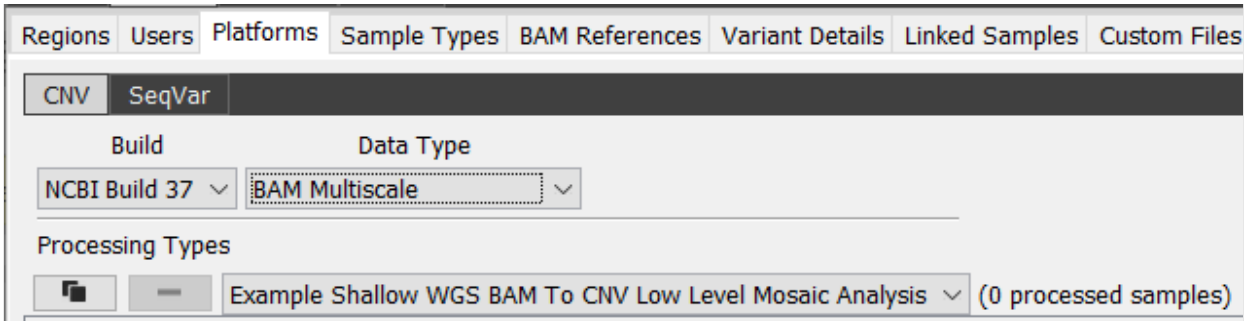


Figure 302. BAM Multiscale.

The provided processing type is Example BAM to CNV analysis. Users can clone this, edit settings, and then associate with a sample type, as shown in **Figure 309**. Note that there is no systematic correction settings section for this processing type as there is for arrays. Systematic correction is performed but it is built into the reference file. See *BAM MultiScale Reference Builder* in the *BAM References* section.

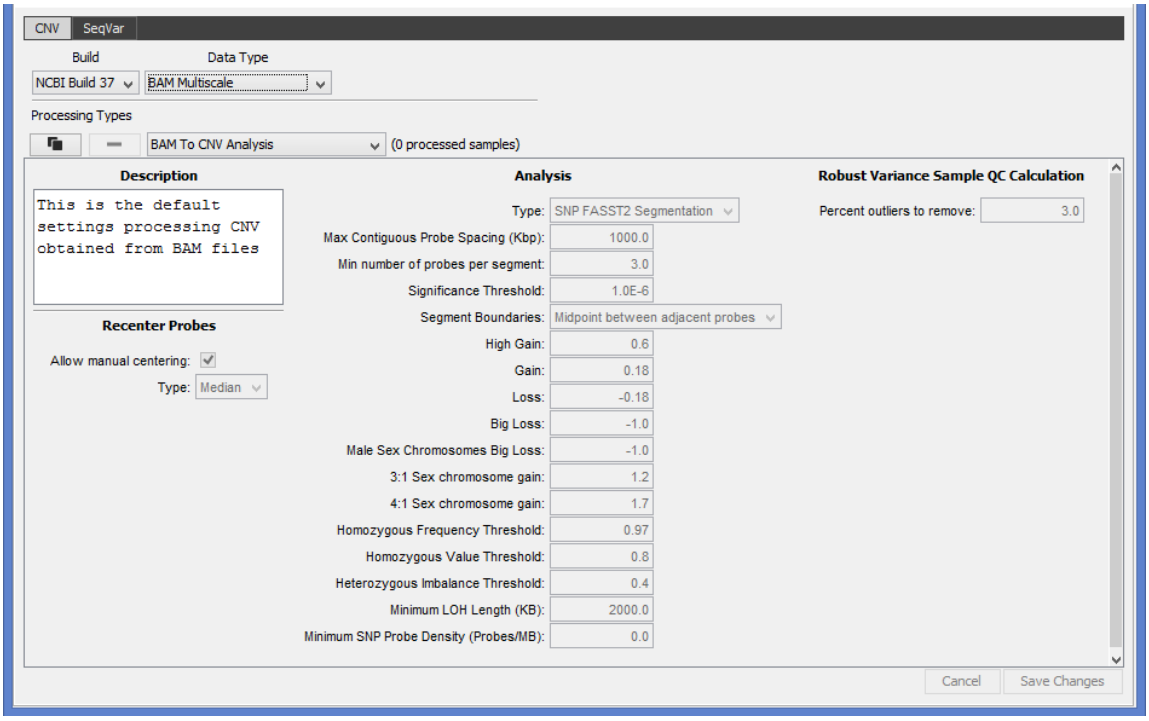


Figure 303. Example BAM to CNV analysis.

BAM References

BAM multiscale processing requires a reference file. The reference files can be uploaded to VIA using the **BAM References** tab. Click on the **+** icon to upload a file. Click on the **–** icon to delete a reference file. See **Figure 310**.

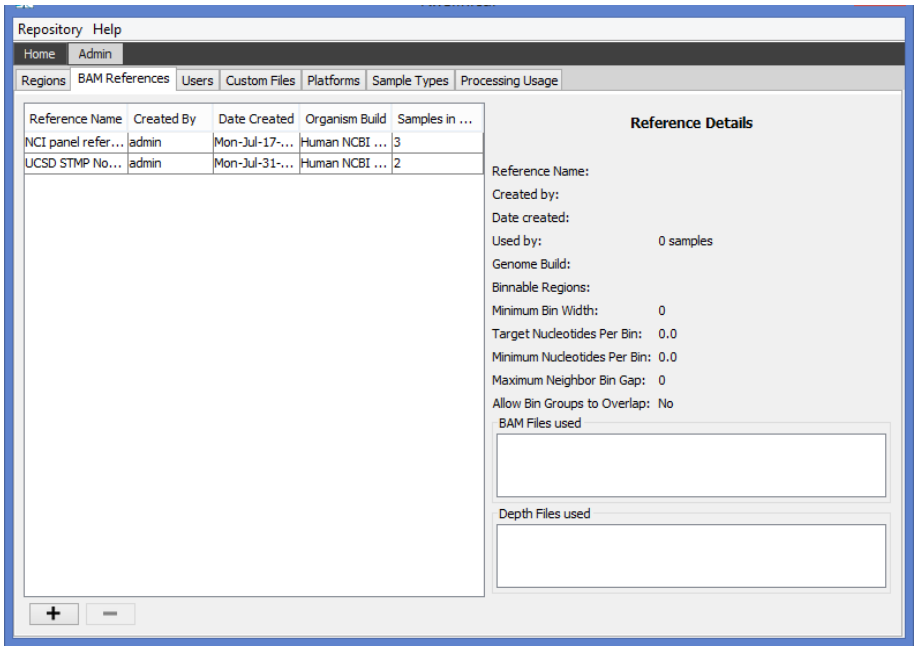


Figure 304. BAM **References** tab.

Highlighting a BAM reference file will display the processing parameters and files used in the right panel, as shown below in **Figure 311**. See *BAM MultiScale Reference Builder* in the *BAM references* section for details on how to create reference files.

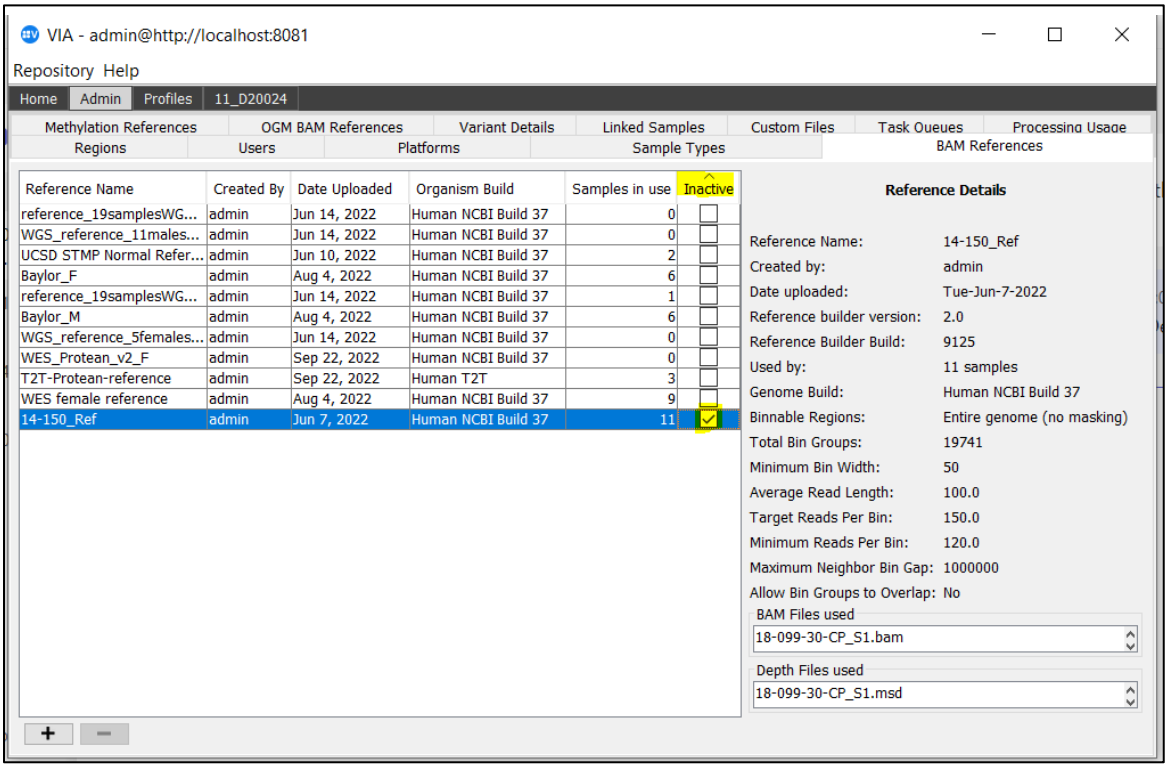


Figure 305. Highlighting a BAM Reference file.

If a newer reference file has been created replacing an older reference, the Admin can designate the older reference file as Inactive by marking the checkbox; thereafter, this reference will no longer be available in the selection during processing.

CNV FROM NGS – BAM SELF-REFERENCE

The **Data Type (BAM Self-Reference)** is another method to estimate copy number from NGS data. The proprietary self-reference algorithm takes the sample BAM, divides the genome into bins according to defined processing values to count reads positionally (bins without sufficient coverage are skipped). It calculates the median read depth per bin and uses this to normalize all bins. The values are converted to log R and systematic correction is applied for any GC effects. The selected segmentation algorithm is then used to call regions of copy number gains and losses.

During the binning process, a masking file is used to define the regions of the genome to bin and is applied to all samples, as seen in **Figure 312**. The masking options are:

- **Mask Undefined (Poly-N) regions** - excludes poly-N regions.
- **Mask Repetitive (Lower-case) regions** – excludes repetitive regions and Poly-N regions.
- **Mask Digital to Analog Converter (DAC) regions** – excludes Poly-N, Lower-case, and DAC regions.

- **Mask Undefined with chrY PAR** – Same as Mask Undefined (Poly-N) regions but including chrY PAR regions.
- **Mask Repetitive with chrY PAR** – Same as Mask Repetitive (Lower-case) regions but also including chrY PAR regions.

DAC Regions are blacklisted regions (anomalous, unstructured, high signal/read counts in NGS experiments) originally created for the ENCODE project (https://www.encodeproject.org/annotations/ENCSR636HFF/). These areas typically surface at spikes near the centromere and telomeres in certain chromosomes.

The bin width parameter needs to be set and there is a tool to help determine what bin width to use for processing. Enter the read depth of the samples and the recommended read depth will be displayed below. If the user wants to use the recommended bin width, click the button and the **Target Bin Width** field will be updated with the recommended bin width. Custom values for the Target Bin Width can be changed manually by clicking in the field and entering the desired value.

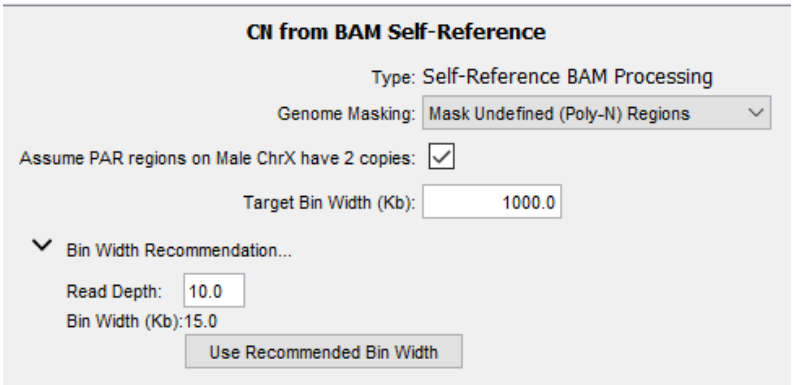


Figure 306. Genome masking file.

In some alignment algorithms, the reads on the pseudo-autosomal regions (PAR) may be equally mapped to both X and Y chromosomes and therefore the PAR region should be treated as an autosomal segment. The parameter Assume PAR regions on Male ChrX have 2 copies when selected will treat PAR regions on ChrX of male samples as autosomes.

Additional parameters can be set for obtaining BAF from BAM files, shown in **Figure 313**. The other parameters are the same as those for BAM MultiScale processing.

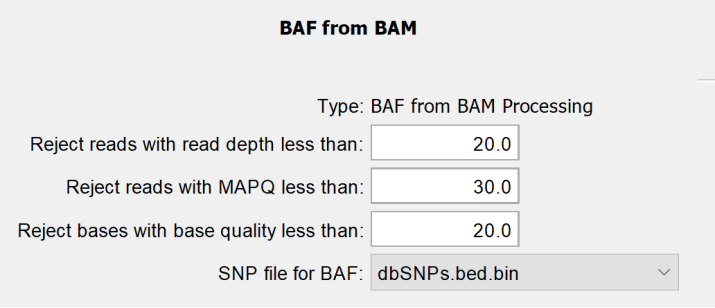


Figure 307. BAF from BAM files.

Capture Bias

For Panels and WES where a capture protocol is used, a **Capture Bias** score is calculated and displayed in the results on the **Home** page and in the **Info.** Window (**Figure 314**). This metric provides an indication of capture efficiency based on read depth distribution in the targeted vs. off-target regions.

Home page:

Status: *Processed* Quality: 0.03 Capture Bias: 0.12 Discarded: 0.00 %AOH: 0.160

Sample Type: Onco Panel - BT Data Type: BAM Multiscale Version 2 Processing Type: Small NGS Panel Mosaic BAM To CNV Analysis

SeqVar Status: *Processed* SeqVar Processing Type: Nirvana

Info. window:

Q/C and Number of Events

Quality:	0.02587074
One copy gain:	34
Two or more copy gain:	2
One copy loss:	33
Two or more copy loss:	0
Capture Bias:	0.12
AOH:	2
%AOH:	0.15991433
%Discarded:	0.0
Total BAM Reads:	68,284,890

Figure 308. Capture Bias score.

Capture Bias values can be positive or negative and values closer to 0 indicate a better capture efficiency. When the score is >1.0 (poor) , the scores will be highlighted in yellow, and a **Re-processing** button will be displayed in the sample Information 4.53 Re-process for capture bias

The user may choose to re-process this sample using an alternative processing that may improve call quality, but it is recommended that such samples be sequenced again for optimal results. Only users with processing privileges will be able to re-process samples. After re-processing, the Info. window displays an inactive button indication that alternative processing has been performed.

Sequence Variants Platform Configuration

The **SeqVar** tab is used to define processing types for sequence variants based on build and data type. Supported file types are NirvanaJSON, VCF, and VCF Nirvana.

Annotated VCF files loaded as Data Type VCF can include annotations from different tools/sources. Ensembl Variant Effect Predictor (VEP) and Nirvana annotations are currently supported in the software. Based on the information available in the VCF file header, the software automatically determines the type of annotations present and applies the appropriate parser. If it cannot find a match, then the annotations cannot be loaded.

The NirvanaJSON file type supports the output from the Nirvana variant annotator when not integrated with VIA.

The Data Type VCF Nirvana supports import of an unannotated VCF file for processing with an installed instance of the Nirvana Annotator to provide a one-step annotation and interpretation workflow. Once loaded, the files are automatically sent to the annotator and the results are displayed as with an annotated VCF or JSON file.

Data Types

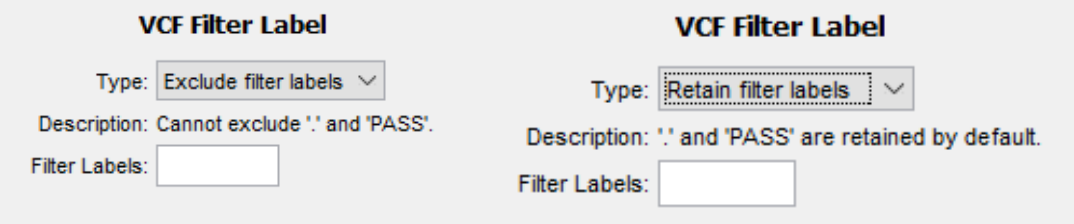
VCF Nirvana: Annotates the input file via the linked Nirvana annotator and then processes the annotated results.

NirvanaJSON: Processes input JSON file preserving the annotated fields.

VCF: Processes input VCF files that were annotated via VEP. Will also allow loading of unannotated VCF, but the data will remain unannotated.

FILTERING EVENTS BASED ON THE FILTER COLUMN IN VCF FILES

The VCF file contains a column Filter which has values that can be used to filter variants. **The VCF Filter Label** parameter in the **Settings** allows exclusion or retention of events matching labels found in the **Filter** column.



Type: This dropdown determines whether matching variants (as per the values specified in Field Labels) will be excluded or retained. Values: Retain filter labels, Exclude filter labels

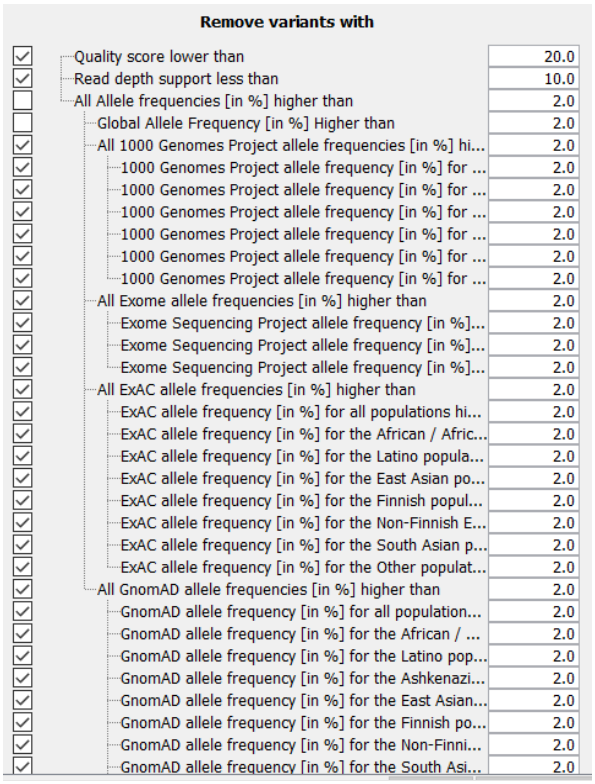
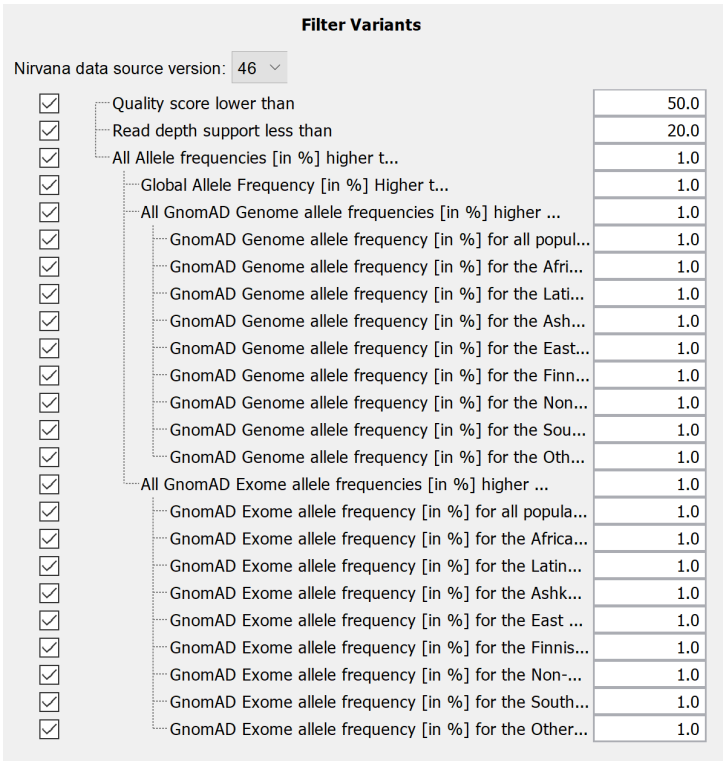
Description: This field is just a note stating that variants with the Labels . and PASS are automatically retained and cannot be excluded. There is no need to specify these in the filter Label if retaining based on filter values.

Filter Labels: Type in comma separated labels (e.g., OffTarget) used in the Filter column of VCF files to use for filtering of variants during processing in VIA. Values here are case sensitive and must match exactly the values in the VCF file.

FILTERING EVENTS BASED ON QUALITY METRICS AND POPULATION FREQUENCIES

The processing parameters for NirvanaJSON and VCF Nirvana are the same as both apply to files annotated using the Nirvana annotator. The processing parameters for VCF are different since annotations are accepted only from .vcf files output from a VEP annotator supported by VIA.

Table 20. Processing parameters.

Settings for VCF:	Settings for VCF Nirvana or NirvanaJSON:
	

VCF Nirvana and NirvanaJSON have additional dropdown fields:

- Nirvana data source version: Selection of the version will result in different filter fields displayed.
- Field for ClinVar label: when marked, will not remove any variant labeled as Pathogenic in ClinVar during processing and subsequent filtering with the UI.

Data source version 46 uses newer cohorts and allele frequencies obtained from gnomAD, both the Global and Exome frequencies. Much of the data from ExAC is incorporated within the new gnomAD frequencies (data source version 46). For detailed information refer to the gnomAD website (About and FAQs). Version 45 contains older data sources and cohorts and should be applied to legacy samples processed via an older Nirvana annotator. Nirvana JSON output files contain data source version numbers and this is the field parsed by VIA to determine compatibility.

Filter usage: The allele frequencies are arranged in a tree-like structure such that selections in the higher nodes apply to all items under that branch. For example, to change all 1000 Genomes Project frequencies to 2.0, make that change in the row highlighted in **Figure 315** and all the population specific groups' frequencies will also be at 2.0. Each individual population can have its own frequency as well – just edit the field for that group.

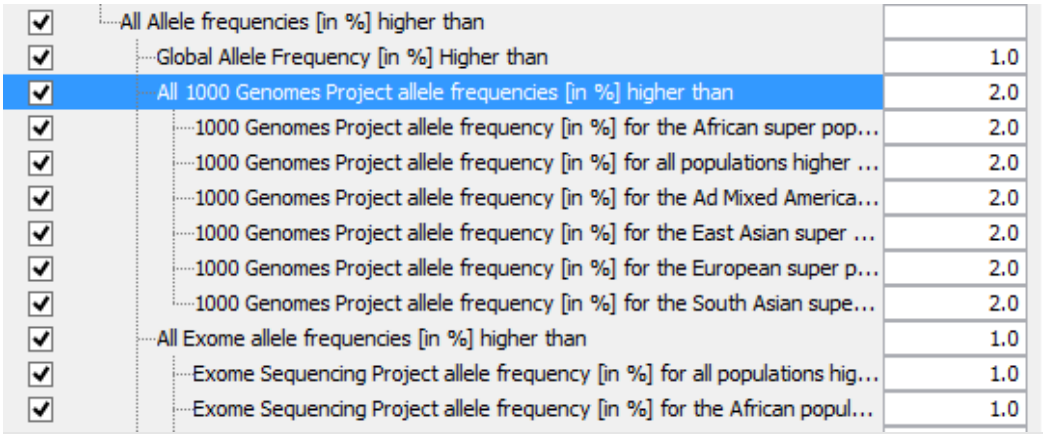


Figure 309. Changing frequencies.

The parameters/filters here are applied to the sample during upload and processing and serve as hard filters for variants prior to the dynamic filters available during case review. The same filters are also available in the sample review interface and the reviewer may apply different filter values to different samples during the review process.

ClinVar Label Filtering: Marking Do NOT remove any variant labeled as Pathogenic or Likely Pathogenic in ClinVar will retain variants marked as such in ClinVar during processing and through further filtering through the Sample Review UI.

PROCESSING TYPES

Creating a New Processing Type: A new processing type is created by copying an existing processing type, renaming it, and adjusting the settings. Example processing types are available for the different platforms. Always create a new processing type (Example processing types are not editable and cannot be applied to a sample type). To create a new processing type:

Go to the Platforms tab:

1. Click on the **CNV** or **SeqVar** tab depending on the platform type (depending on license type, only CNV may be available).
2. In the **Data Type** dropdown, for CNV, select the build, data type, manufacturer, and assay name for arrays; select the build and BAM MultiScale or BAM Self-Reference for estimating copy number from NGS. For SeqVar, select the build and data type.
3. Next, select a processing type from the dropdown.
4. Now click on the **Copy this processing type** button, which prompts for a name for the new processing type. Once the new name is entered, it will be displayed in the dropdown under the **Processing Types** section.
5. Modify the settings. Once all changes have been made to the settings, click on the **Save Changes** button on the bottom right of the window to save this processing type. Now this processing type is available to be applied when defining a sample type.

Deleting a Processing Type: A processing type can only be deleted if there are no samples in the database that have used that processing type and if it is not an example processing type. To the right of the **Processing Type** dropdown, the number of samples processed with these settings is indicated. If there is even a single sample processed with a specific processing type, the processing type cannot be deleted.

In **Figure 316**, the processing type is not an example one and there are no samples associated with this processing type so the **Delete** button is not grayed out and the processing type can be deleted.

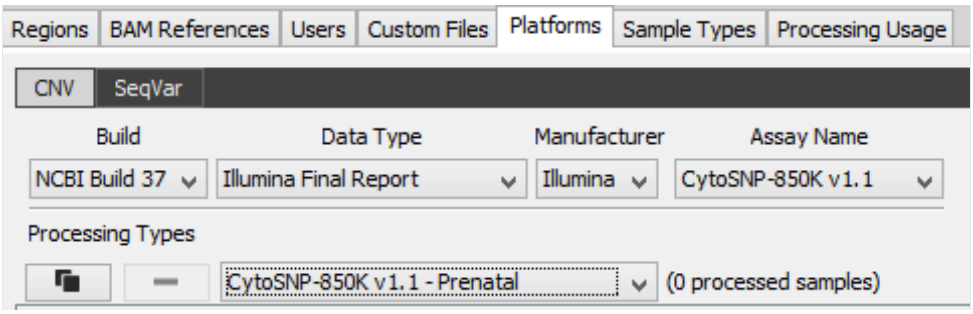


Figure 310. The processing type can be deleted.

When attempting to delete a processing type, an alert asking for confirmation will be presented, as shown in **Figure 317**.

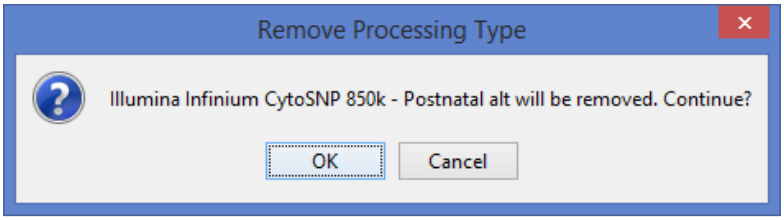


Figure 311. Confirmation.

Editing a Processing Type: The settings for a processing type other than an example type can be changed by selecting the processing type and then editing the values for the settings. Once all changes have been made, click on the **Save Changes** button on the bottom right of the window. To alter the settings, simply click on **Cancel** to revert to the previously saved values.

If the processing type already has some samples that have been processed, attempting to change the settings will bring up an alert, shown in **Figure 318**, stating that the samples will need to be re-processed using the new settings.

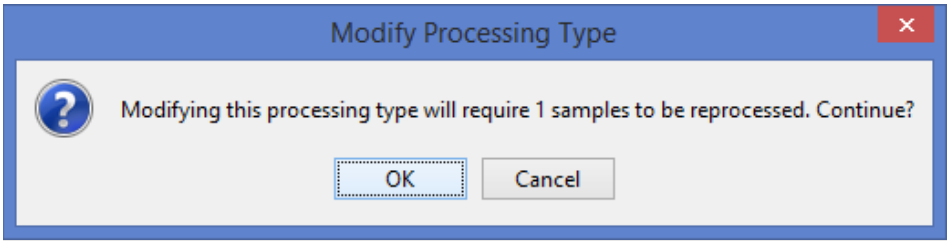


Figure 312. Reprocessing alert.

Click **OK** to continue with editing settings or **Cancel** to revert to the original settings. After all settings changes are made, click **Save Changes**. Now all processed samples (with the exceptions mentioned above) using this processing type will become unprocessed. These samples will need to be processed from the **Home** page.

Only samples that are not locked and not in the review workflow stage can be re-processed. If a sample is locked or has been reviewed, a pop-up box appears that tells you the processing type cannot be modified, as seen in **Figure 319**.

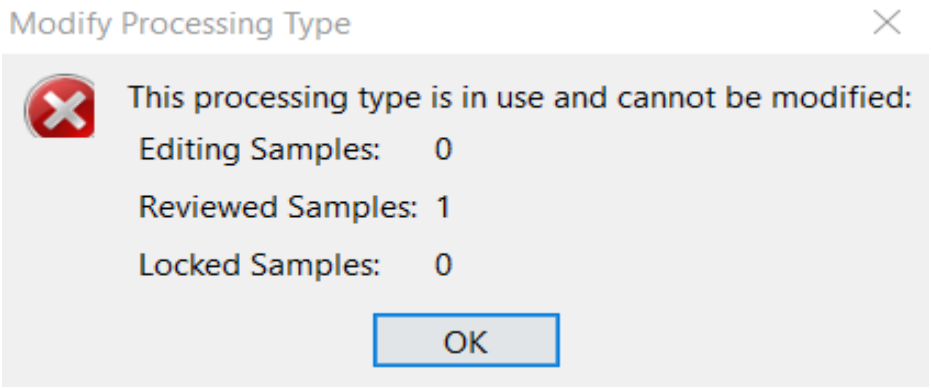


Figure 313. Cannot reprocess alert.

Sample Types: In the **Sample Types** tab, the Administrator creates the available sample types by specifying allowed sample attributes, classification categories, decision trees, default sample view preferences, reports, and a processing protocol.

A sample type is the admin-controlled composite of platform technology, processing settings, classification, reporting, and other configurable elements for the analysis of samples according to a particular application. There is no limit to the number of sample types that can be created. There are many uses for creating differing sample types, for example, sub-categorization associated samples in the VIA database.

Tools across the top allow creation, deletion, editing and copying of sample types. The dropdown field lists the sample type and next to it in parentheses is the number of samples of this type in the database as well as the Sample class. In **Figure 320**, one can see that there are no samples of the sample type Affymetrix Cytoscan HD - Cancer.

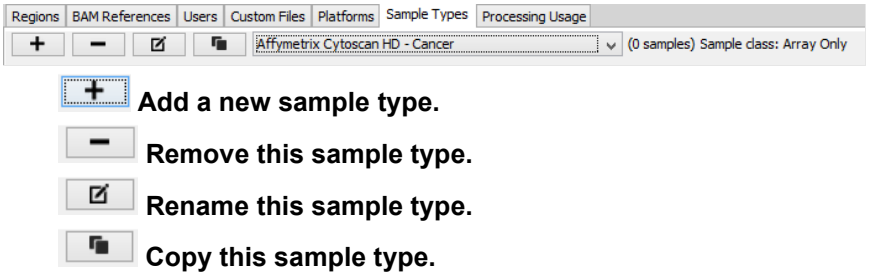


Figure 314. Sample Type tools.

Based on the sample class, parameters for one or both modalities (CNV/SeqVar) may be specified. In **Figure 321,** the sample type has both copy number and sequence variants, therefore both boxes are checked for this sample type.

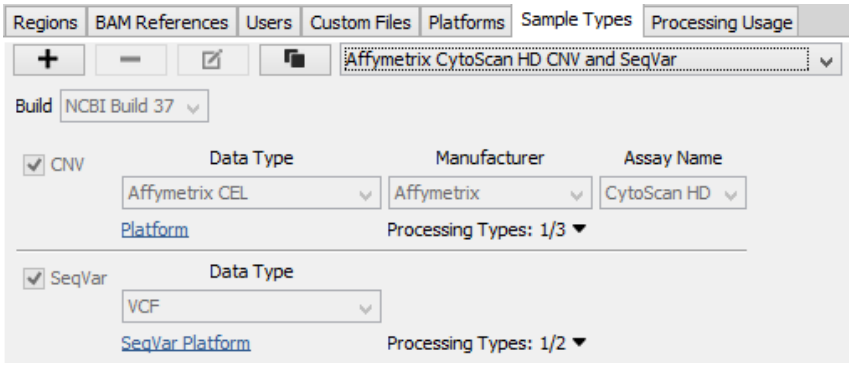


Figure 315. CNV and SeqVar parameters.

NOTE: More than one processing type may be associated with each sample type. During sample upload and processing the user will select which of the associated processing types to apply.

Sample Type Configuration

Creating a new Sample Type: A sample type belongs to one of several sample classes. A sample class is defined by different factors which depend on the modalities (CNV, SeqVar, or both) associated with the sample type.

Sample Class: The sample classes are Array Only, NGS and Array, GxA-Cyto, Low-Res WGS, Methylation, OGM and NGS, OGM Only. When creating a new sample type, a sample class must be selected. In addition, the modality (CNV and/or SeqVar) will need to be selected. In other words, if the data only has CNV information, users will check the **CNV** box. If the data only has sequence variant information, check the **SeqVar** box. If the data has both copy number and sequence variant information, then check both boxes.

Test Type: A test type defines certain module/features available to the sample type allowing display of customized fields in various areas such as the KB and Variants Details to capture and display relevant information for a specific type of sample. A test type value is not required (can be blank). If left blank, the sample is treated as constitutional, and some functions may not be available with the KB features.

Mitochondrial Chromosome inclusion: The **Include chrM** checkbox specifies that the mitochondrial chromosome in the sample should be included in analysis.

CNV modality: Click on the **Add a New Sample Type** button and enter the sample type name in the pop-up window. Select the sample class from the dropdown, seen in **Figure 322**. Sample classes available here are based on those allocated to the license with other options grayed out.

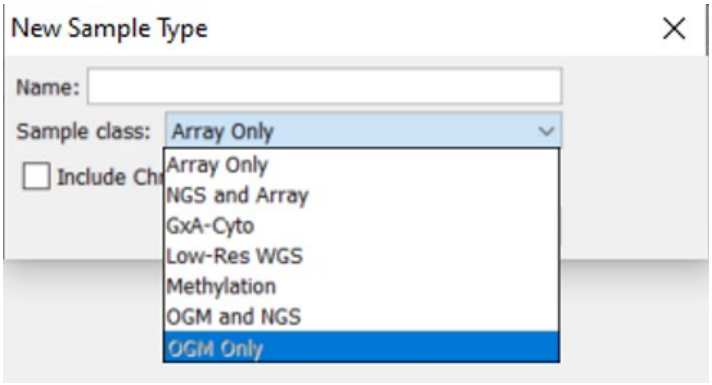


Figure 316. Select sample class.

Next make sure only the **CNV** box is checked off and select the **Genome Build**, **Data Type**, **Manufacturer**, and **Assay Name** from the dropdown fields. If expected values are not seen here, review the section on adding a new data type (Creating a New Data Type) or contact support@bionano.com.

NOTE: For the GxA-Cyto sample types, CNV is checked by default and cannot be unchecked.

From the **Processing Types** dropdown menu, shown in **Figure 323**, select the processing type desired for this sample type. Processing types are the parameters and settings that were defined in the **Platforms** tab for the data type.

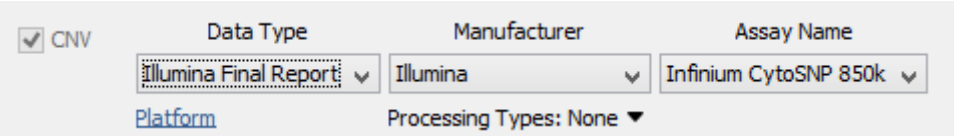


Figure 317. Select the processing type.

To improve awareness of the sample quality, the Admin can choose warning thresholds, shown in **Figure 324**, which will color code the QC values on the **Home** page. If the Quality score of the sample exceeds the threshold, the score is highlighted as such:

- exceeds the Warn threshold -> highlighted in yellow.
- exceeds the Fail threshold -> highlighted in red.

Figure 318. Quality thresholds.

The CN quality score is calculated after removing outliers. The percentage outliers to remove is set in the **Processing** settings in the **Robust Variance Sample QC Calculation** section and is specified as a percentage. If 0.2 is the specified value, 0.1% of outlier probes from the bottom of the spectrum will be removed and 0.1% will be removed from the top.

SeqVar modality is only available to sample classes with sequence variant analysis permitted, such as NGS and Array.

Select the **SeqVar** box, as shown in **Figure 325**.

NOTE: For GxA-Cyto sample types, the **CNV** box cannot be unchecked. CNV processing can also occur in parallel, if desired. Select the desired processing types available for this sample type and click **Save Changes**.

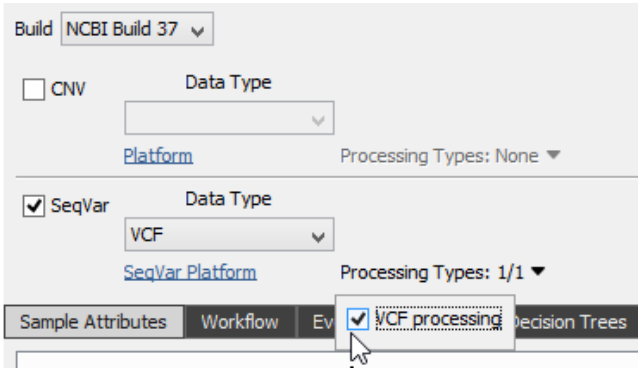


Figure 319. SeqVar box.

Copy Number from NGS: To create a sample type for obtaining copy number from NGS data, select **NGS** and **Array** in the **Sample Class** dropdown and provide a name for the new sample type, as in **Figure 261** above.

Select the **CNV** box and select **BAM Multiscale** from the **Data Type** dropdown. Select the processing types to associate with this sample type.

Select the **SeqVar** box and select the data type (**NirvanaJSON**, **VCF** or **VCF Nirvana**) based on the type of sequence variant files available for this type of sample. Select processing type(s) from the dropdown to associate with this sample type. Click **Save Changes**. See **Figure 326**.

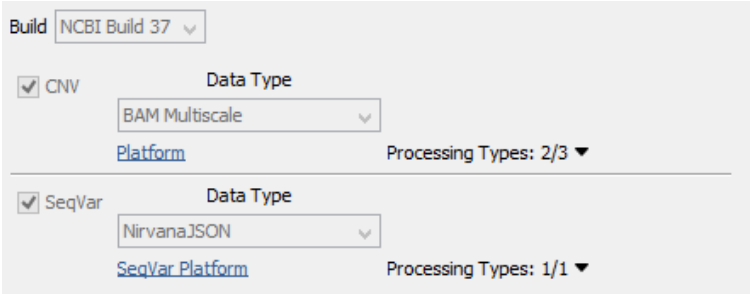


Figure 320. An example sample type for obtaining CNV from NGS data.

Relationship between Processing Type and Sample Type: At the basic level, both a sample type and processing type are defined by the same components, but a sample type has one or both modalities (CNV and/or

SeqVar) whereas a processing type has the same components for a single modality (CNV or SeqVar). The BAM Multiscale 1 sample type, shown in **Figure 327**, has both CNV and sequence variants so both components below are available:

1. For CNV:
   - Genome Build (e.g., NCBI Build 37)
   - Data Type (e.g., Illumina Final Report)
   - Manufacturer (e.g., Illumina)
   - Assay Name/Array (e.g., Illumina 850K)

   For arrays, all four of the above are available.
   For copy numbers derived from NGS, only Build and Data Type are available.
2. For sequence variants:
   - Genome Build (e.g., NCBI Build 37)
   - Data Type (e.g., VCF, NirvanaJSON)

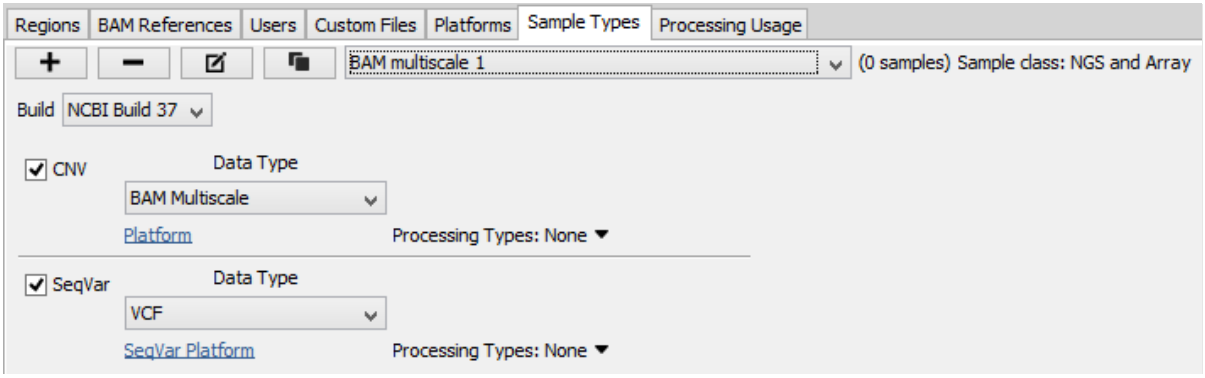


Figure 321. BAM Multiscale 1.

A Platform is also defined by the same set, displayed in **Figures 328** and **329**:

- Genome Build (e.g NCBI Build 37)
- Data Type (e.g Illumina Final Report)
- Manufacturer (e.g Illumina)
- Assay Name/Array (e.g., Illumina 850K)

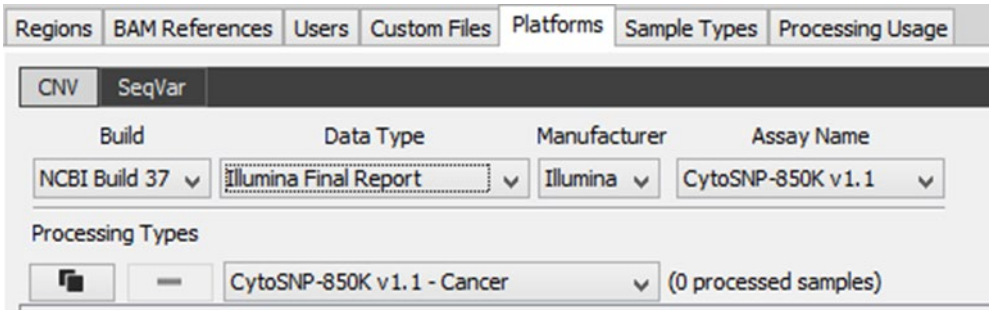


Figure 322. An array platform.

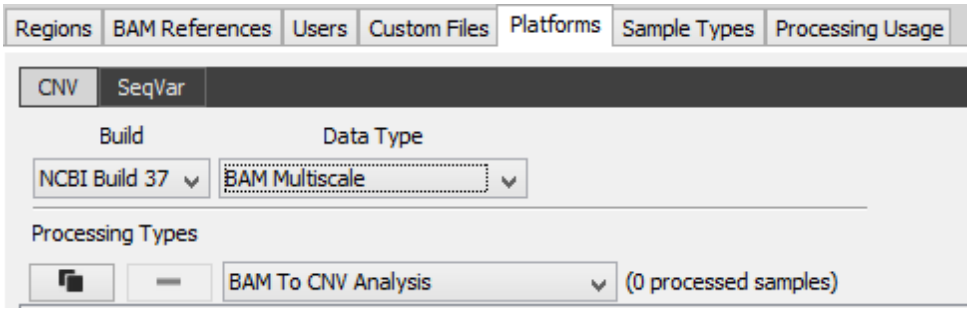


Figure 323. NGS platform (BAM Multiscale) for CNV.

The VIA Administrator can create one or more processing types for each platform. Each processing type has its own set of parameters (e.g., gain/loss cut offs, systematic correction file, ability to re-center probes). For example, if there are two different systematic correction files for the array type Illumina 850K, two different processing types can be created, each using a different systematic correction file.

The VIA Administrator can allow one or more processing types to be used when processing samples of a given sample type by selecting those shown in **Figure 330**.

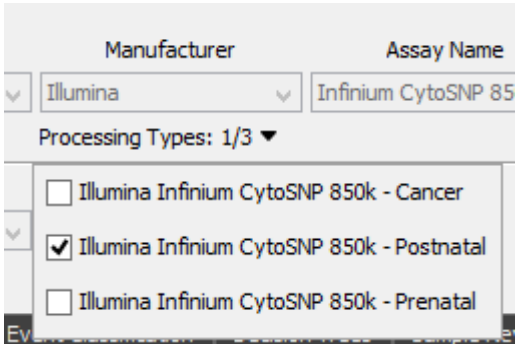


Figure 324. Processing types.

Processing Types: Parameters are specified in the **Platforms** tab, as shown in **Figure 331**. A single processing type is applied to a CNV modality or a SeqVar modality as each component has different parameters for processing. Multiple processing types may be associated with each modality of a sample type. Processing types available to a sample type are listed in the dropdown field. Only those processing types selected will be actively available for processing samples.

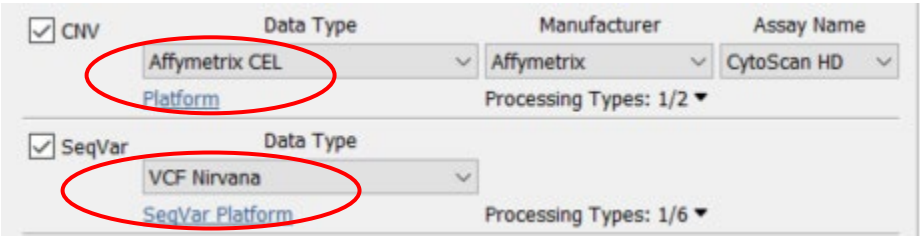


Figure 325. Each sample type has associated processing types for each modality.

The user will select the processing type to use in the **Process Samples** window, shown in **Figure 332**.

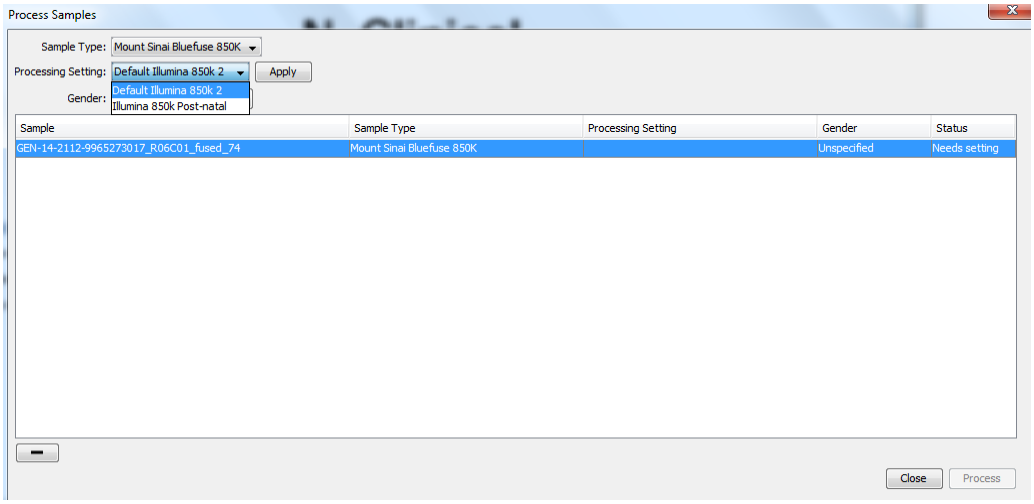


Figure 326. The **Process Samples** window.

Removing a Sample Type: Select the sample type name via the dropdown and click on this tool to remove the sample type from the system. Note that this tool is active for a specific sample type only if there are no samples of that type in the database. If there is even a single sample of that type in the database already, the icon will be grayed out and the sample type cannot be deleted.

Renaming a Sample Type: Select the sample type name via the dropdown and click on this tool to rename the sample type. A pop-up window will open with the existing name in the **Edit** field. Note that this tool is active for a specific sample type only if there are no samples of that type in the database. If there is even a single sample of that type in the database already, the icon will be grayed out and the sample type cannot be renamed.

Creating a New Sample Type by Copying an Existing Sample Type: To create a new sample type like one that already exists, select the name via the dropdown and click on the **Copy this sample type** tool. A pop-up window will ask for a new name. This new sample type will have all the same properties as the one that was copied but to change them for the new sample type, edit the properties and click **Save Changes** on the bottom right of the window.

Sample Attributes: Here, the Administrator defines annotation information associated with samples as sample attributes. The administrator defines a sample attribute name and by default, each field will be a free-form text option. However, if the attribute has a fixed set of inputs, these can be listed, allowing the Administrator to constrain the different possible input values and allow filtering based on attributes when querying for samples.

The Administrator inputs the potential values (Labels) available to users. For example, the attribute gender can be constrained to only allow the following values: Unspecified, Male, Female. Constraining this prevents inconsistencies in terminologies among different users, preventing users from inputting M, male, or other variations to specify the gender as male.

The attributes are defined on a per sample type basis so each sample type can have different attributes associated with it. **Attributes Display Name**, **Gender**, **Phenotypes**, **Linked Sample Relationship**, and **Linked Sample ID** are default attributes for all sample types and cannot be deleted. Certain sample types such as Illumina data have an additional default attribute Filename which cannot be deleted.

Figure 333 shows the Affymetrix CytoScan HD - Postnatal sample type with the default attributes and the default Labels for the Gender attribute.

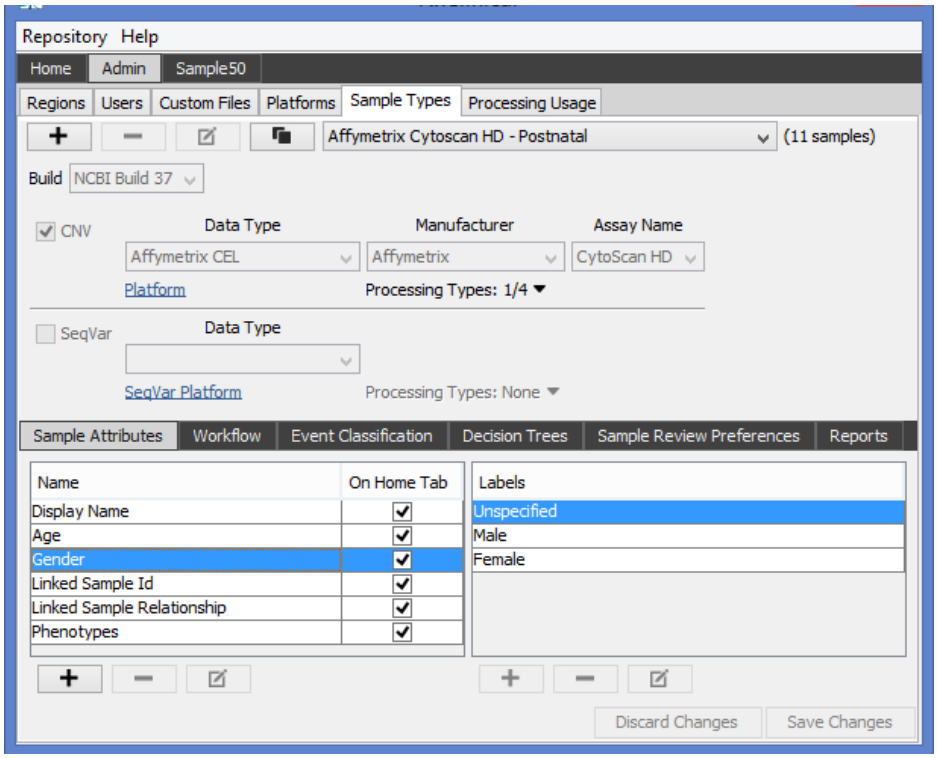
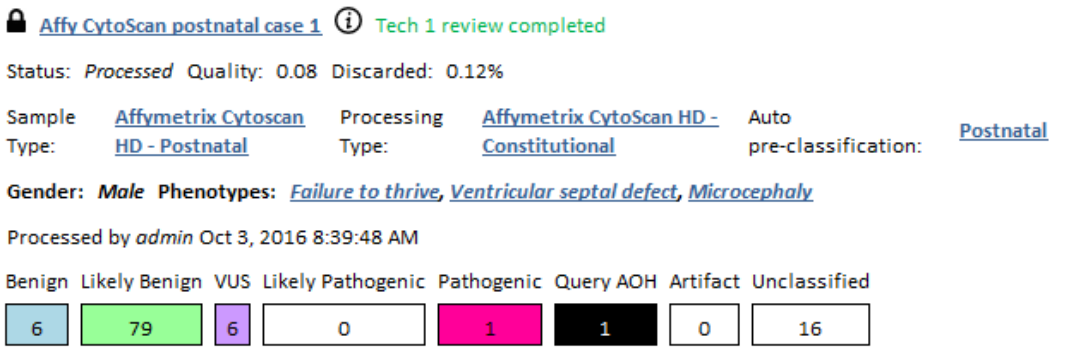


Figure 327. Sample Attribute is Gender.

The **On Home Tab** checkbox field controls the display of the attribute in the query results on the page. If the checkbox is marked, the attribute will be displayed (if it contains a value) on the **Home** page query results. In the query result below, **Gender** is displayed as it was checked off for the Affymetrix CytoScan HD – Postnatal sample type. Even though **Linked Sample ID** is checked off, it is not displayed as this sample has no value for that attribute.



The option to hide or display attributes is particularly useful for some sample types that may have an abundance of associated physical sample run information not necessary to display during queries. The sample type shown in **Figure 334** has a good deal of specific run fields associated as attributes but only five (checked off) will be displayed in the query results on the **Home** page.

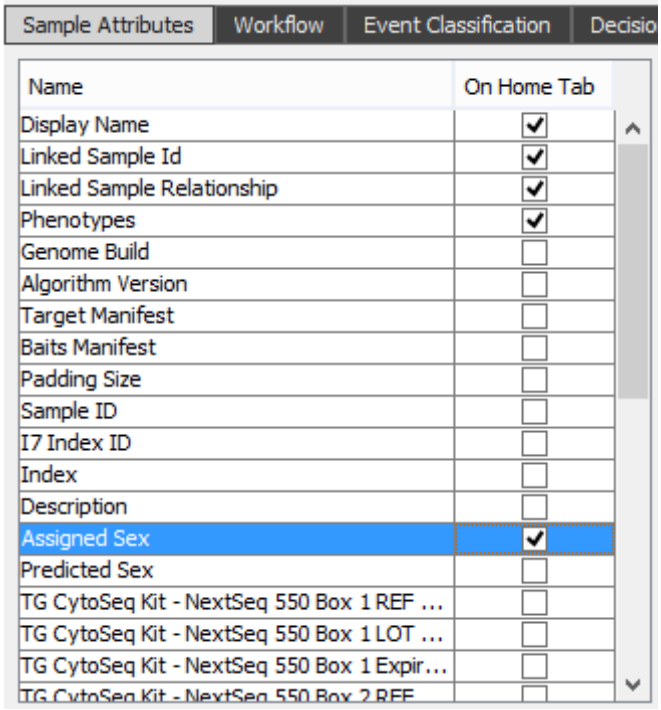


Figure 328. Many specific run fields.

Affymetrix OSCHP data types: When a new sample type of the Affymetrix OSCHP data type is created, additional default attributes are associated with this data type. The software also loads some of the QC values from the OSCHP file and displays them in the **Information** window. In **Figure 335**, one can see these additional attributes listed in the **Sample Attributes** tab.

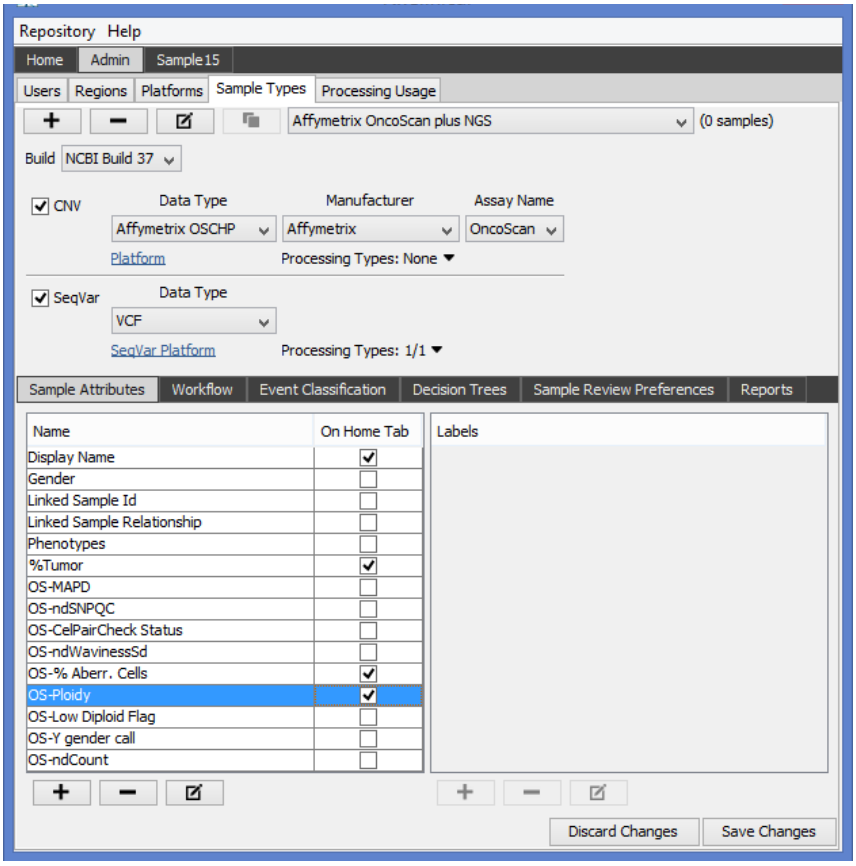


Figure 329. QC values not displayed.

Date attributes: Attribute values when entered in a specific format will be treated as dates. VIA assumes attributes with values in the format ####-##-## are dates in the date format yyyy-mm-dd. One could have an attribute **Date of Birth** with values such as 1968-08-20 and the system considers this as August 20, 1968. Such values can then be used to search samples by date (e.g., looking for all samples with date of birth prior to January 1, 1980).

DEFINING SAMPLE TYPE PROPERTIES

To create a new attribute, click on the **Add a New Sample Attribute** button and enter a name for the attribute in the new row in the **Attributes** table. Check off the box to display the attribute in query results on the **Home** page, shown in **Figure 336**.

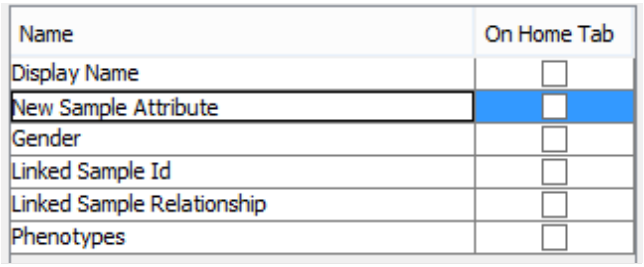


Figure 330. Creating a new attribute.

To add labels, click on the **Add a New Label** button under the **Labels** table. A new row will be added and a name for the label can be entered, seen in **Figure 337**.

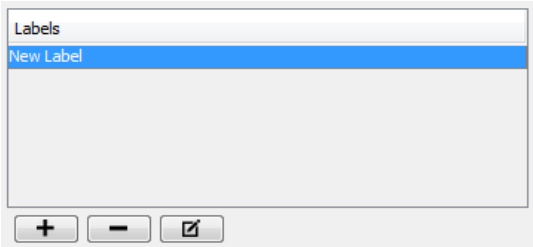


Figure 331. Creating a new label.

Use the respective **Edit** tool to change an attribute or label name. Once all attributes and labels have been added, click on **Save Changes** to save the information for this sample type. To remove labels or attributes, select the row(s) and click on the **Remove Selected** button . An attribute can only be deleted or edited if no sample has a value for that attribute.

Linked Samples Tab: The Admin can further customize the labels for linked samples to best represent the lab's workflow or local language. The **Linked Samples** Tab has a list of labels used for linked sample relationship and the associated relationship meaning. Some family relationships need to have a standard meaning that the software can understand for calculations such as trio quality check, parent of origin, and recessive inheritance filtering. The relationship meaning values are selected via a dropdown and restricted to the following: Father, Mother, Proband, Sibling. For example, the Admin could create a new Label called Dad which would have the meaning Father, as seen in **Figure 338**.

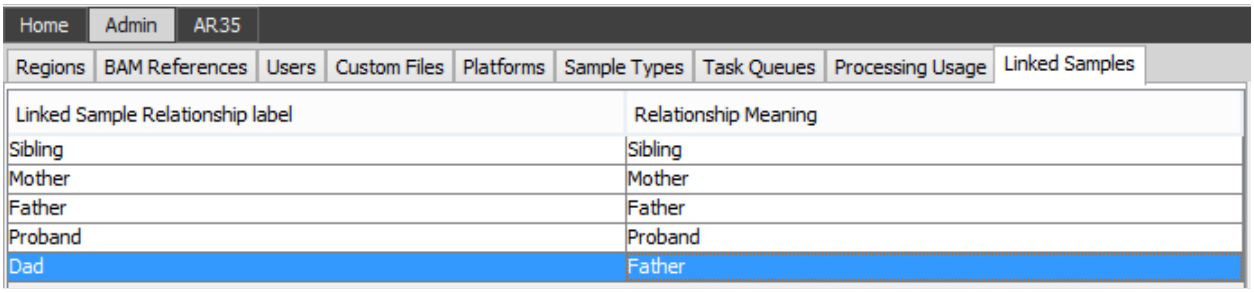


Figure 332. Dad means Father.

To add a new label, click the **+** button and enter a label. Use the dropdown to select a relationship meaning and click **Save Changes**. Now the label Dad can be used in the section, as seen in **Figure 339**.

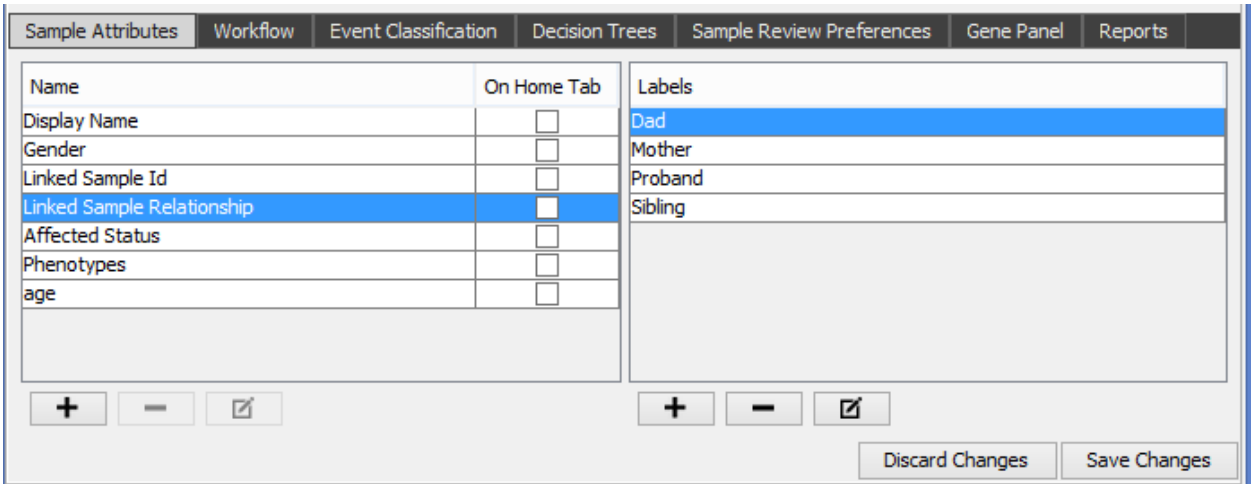


Figure 333. Label has changed in Sample Attributes.

The Phenotypes attribute: The Phenotypes attribute is a special attribute that allows association of HPO terms with the sample. This is specific to each individual sample and therefore added during or after sample upload. See the *Creating a Sample Type, Sample Loading and Processing* section for details on associating phenotypes.

The VIA Administrator can define different workflow stages, as shown in **Figure 340**. Available tools:

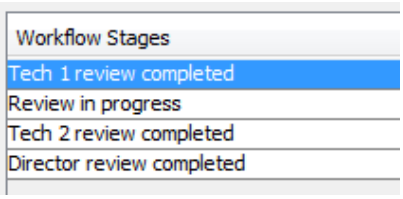
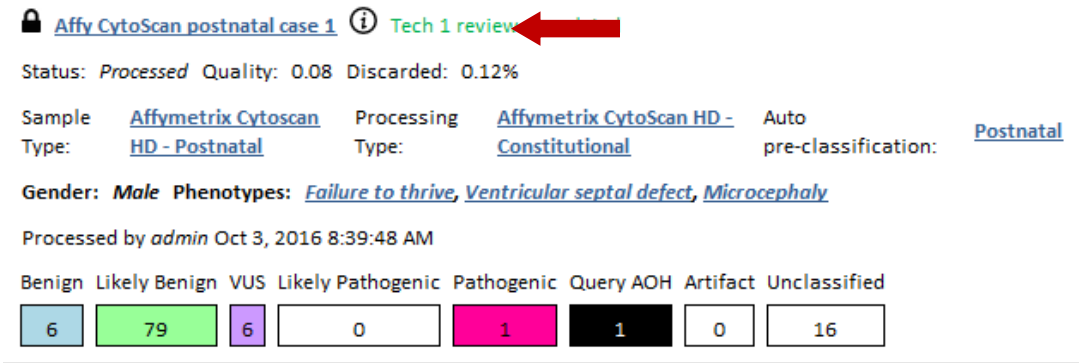


Figure 334. Different stages of workflow.

Click the **+** button to add a workflow stage. Highlight a row and click the **–** button to delete the stage. Highlight a row, click the **Edit** button, and edit the text to change the name of the workflow stage.

The workflow stages allow the Admin to see where in the process the sample is (e.g., has the director reviewed it?). During the review process in the editing mode, a reviewer can change the status. For example, if the first technician has finished reviewing, then the value can be set to Tech 1 review completed. This value is then displayed (in green text) with the list of samples in the sample search interface:



The status can only be changed in the sample edit mode. Click on the start editing tool and then the workflow stages will be visible in the dropdown, as shown in **Figure 341**. The stage can then be selected from the dropdown menu.

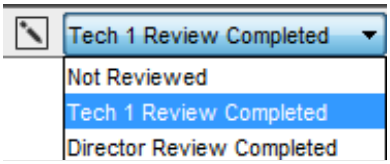


Figure 335. Workflow stages.

Event Classification: This section defines the different classification values that are available with an associated color for graphical representation. Also, the decision tree to use for the automated classification (pre-classification) is selected here. Use the **Add**, **Remove**, and **Edit** tools to make changes to the available classification values, seen in **Figure 342**.

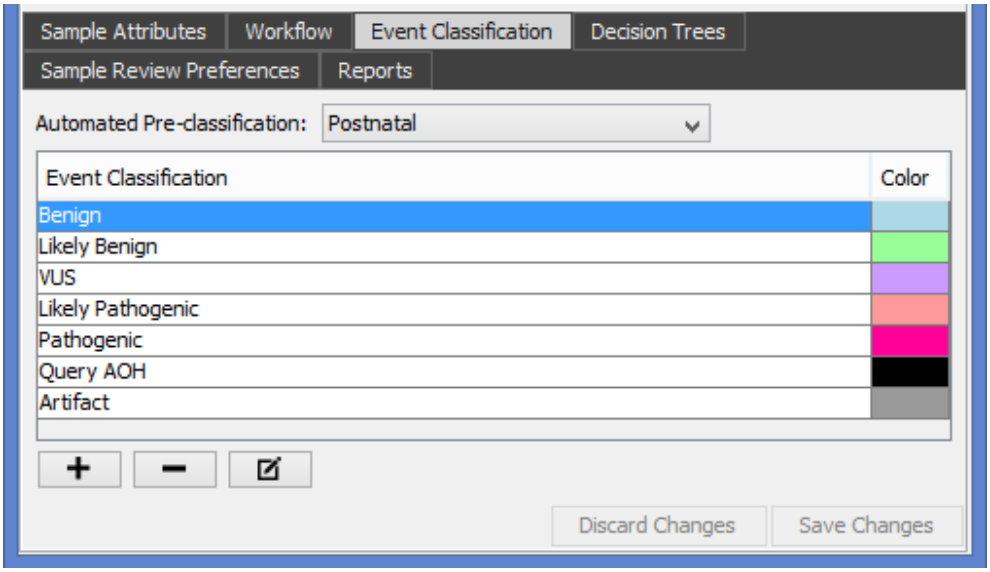


Figure 336. Event Classification.

Classified events displayed in the tracks are color coded based on the colors selected here. To change the color associated with the classfication, select a classification value (click on the row), then click the **Edit** button and a window pops up, as seen in **Figure 343**.

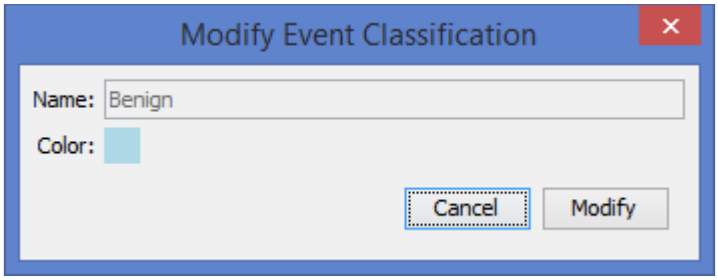


Figure 337. Window for color.

Click the color box to select a new color from the pop up color chooser, shown in **Figure 344**, and then click **OK**.

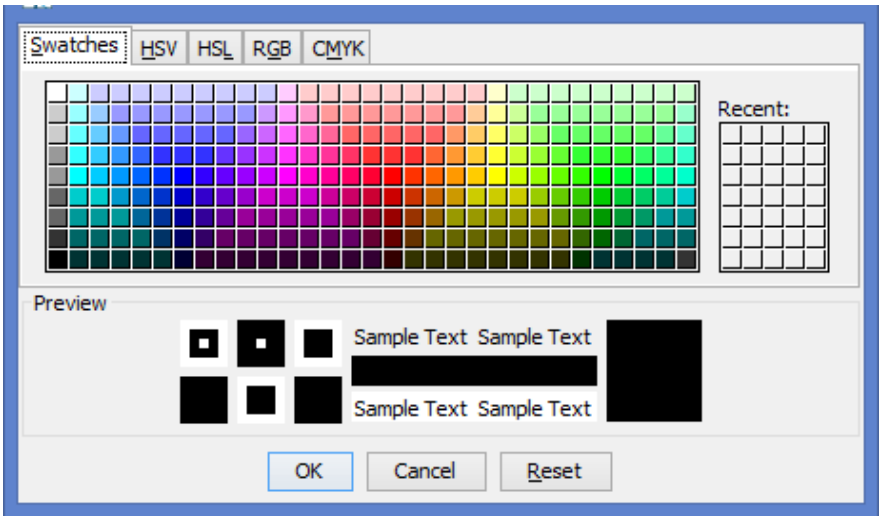


Figure 338. Pop-up color chooser.

Finally, click Modify in the **Event Classification** window to save the new color.

Figure 345 shows a CN loss displayed on the browser in the **CNVevent** track; the pale green colored bar under the red loss indicates the Likely Benign classification.

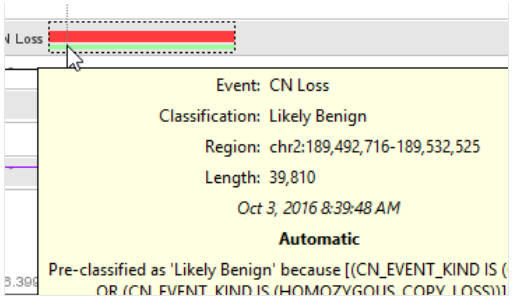


Figure 339. Likely Benign.

The decision tree to apply to the sample type for automated pre-classification can be selected here as well via the dropdown field, shown in **Figure 346** in the red rectangle. This pre-classification engine is run during initial sample processing and can be run again when any changes are made to events for a sample. When an event is manually altered, a pop-up alert will ask the user if automated pre-classification should be run, after each manual

change. If a user is making many changes, the prefererence may not be to have this pop-up appear constantly. To prevent such pop-ups, uncheck the box **Enable Automated Pre-classification** on manual edits.

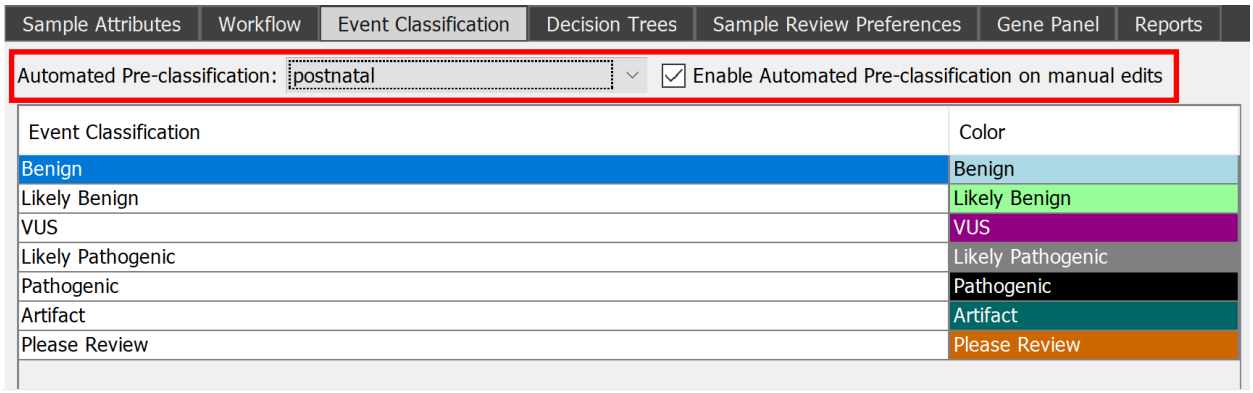


Figure 340. Drop-down for the selected Decision Tree.

When a sample that has a decision tree associated with the sample type is processed, the decision tree logic runs and pre-classifies events based on the rules. The logic that triggers the classification is then recorded in the **Notes** section. **Figure 347** displays an example of a simple decision tree called Test 1 and a more complex one is shown in **Figure 348**.

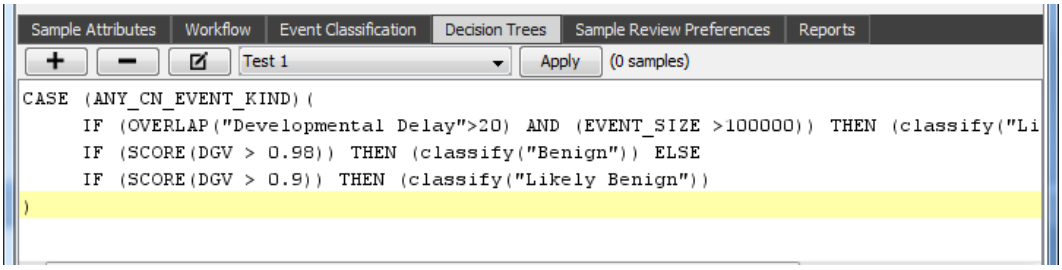


Figure 341. A simple decision tree.

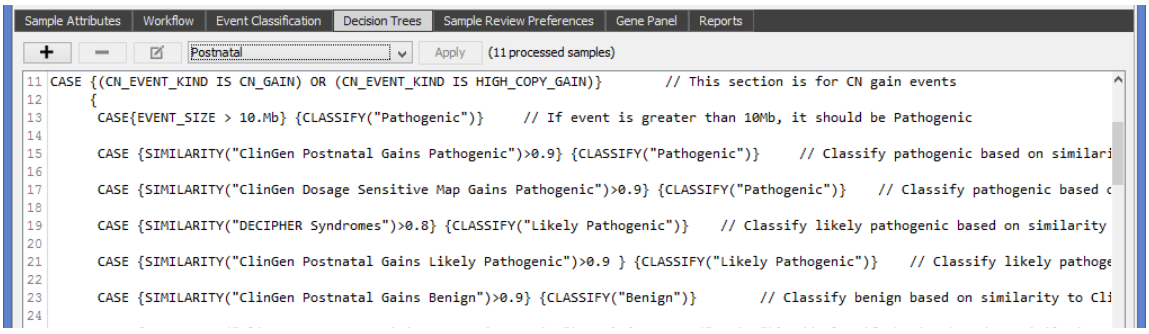


Figure 342. A more complex decision tree.

Decision Tree: The rules applied for automatic pre-classification of events are based on the defined criteria outlined using a specified syntax. Specific keywords and syntax are used to create the decision tree rules. Failure to follow the syntax carefully will result in errors and the automated classification may not work (for example, parentheses and curly brackets match). The functions used for the decision tree rules are case-sensitive so attention must be given to this as well.

Detailed guidance for the construction of the pre-classification decision tree (DT) is provided separately. Bionano support team will assist with generating the DT language and scripts to mirror the logic used in the lab for the interpretation process.

Please refer to the VIA pre-classification syntax information found in the *VIA Theory of Operations* (CG-00042) for details on the functions and how to write the decision tree script.

Sample Review Preferences: The **Preferences** tab allows setting of the default displays for table columns, tracks, and filters for a sample type. The end user can change these for a view of a specific sample or for all samples from the individual sample view. Different display preferences can be set for the various sample types. Select the sample type from the dropdown at the top and then set the default column, tracks, and filters settings, as seen in **Figure 349**.

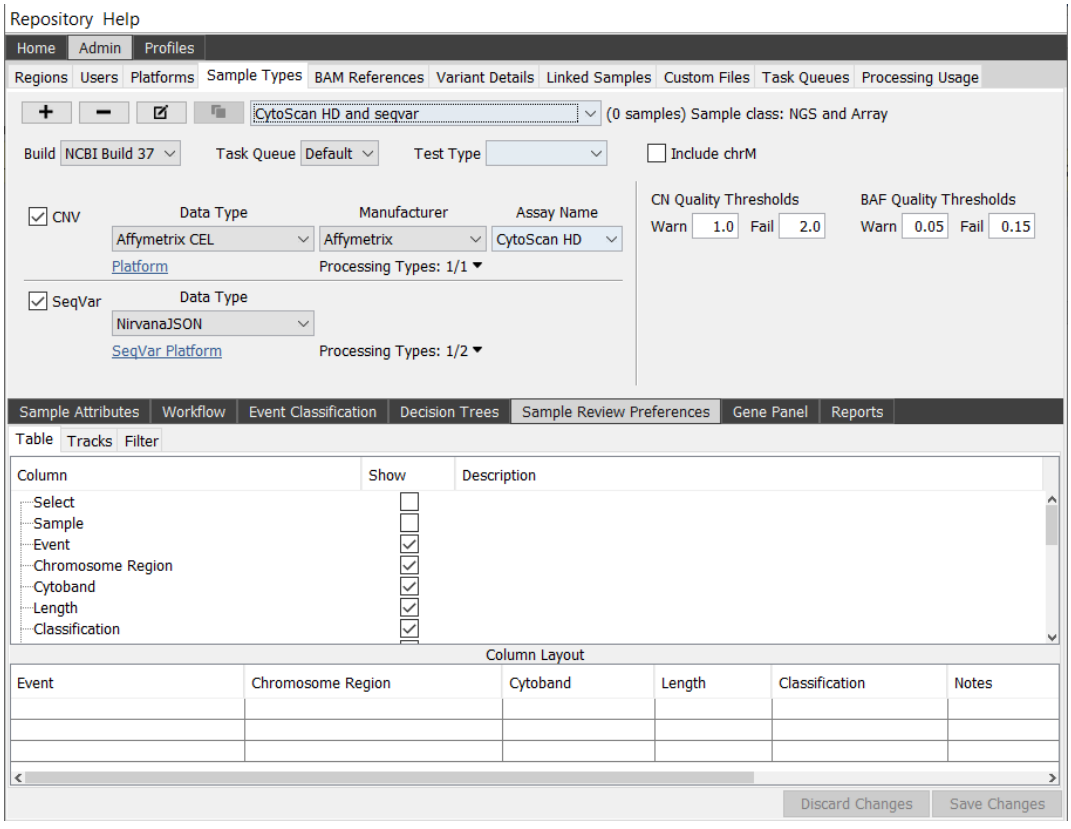


Figure 343. Setting the default column, tracks and filters.

The Table tab (see **Figure 350**)**:** The Administrator can select the columns (by checking off the boxes) to be displayed when a user views a sample. Columns specific to certain types of variants are grouped together in two folders: **CN and Allelic Events**, and **Sequence Variant Events**. Within the **Sequence Variant Events** folder is another one entitled **Transcripts** which lists columns specific for transcript details.

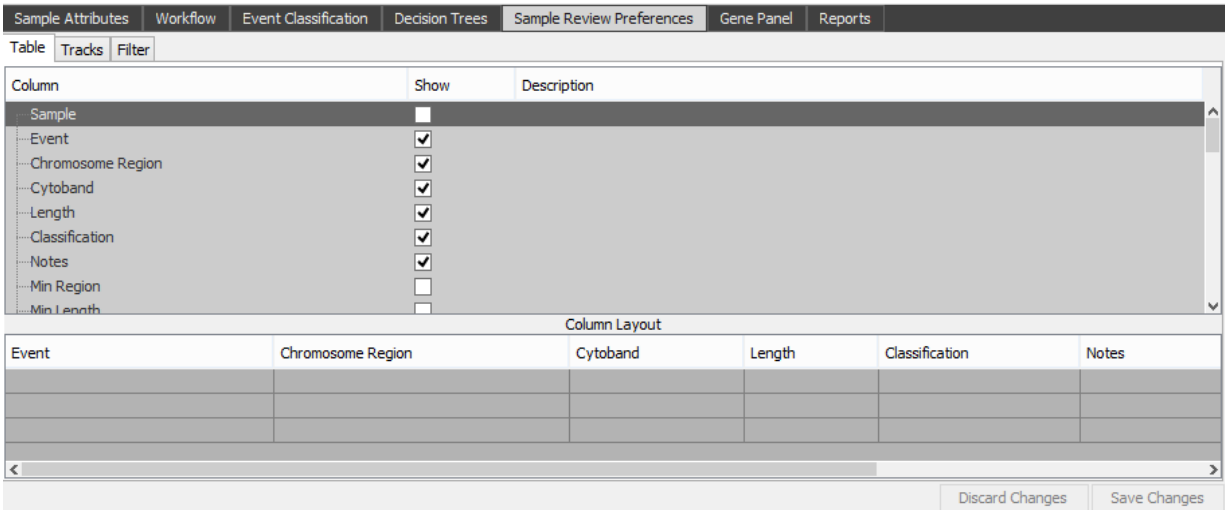


Figure 344. The **Table** tab.

Scrolling down to the bottom of the list of columns, Regions and Decision Trees are displayed, as seen in **Figure 351**.

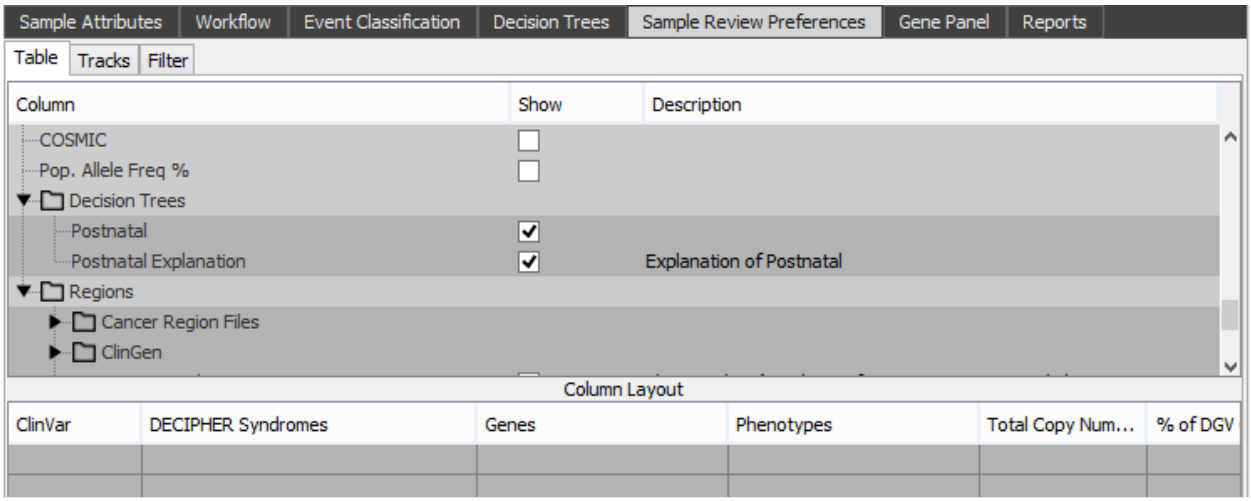


Figure 345. Regions and Decision Trees.

The decision trees folder shows the different decision trees for the sample type. Check boxes to display these columns in the sample review. Note that the decision trees columns will **only be displayed for Administrator user accounts**. These columns are displayed for the Administrator so that different decision trees can be applied to see how they perform and choose the best one to use for a specific sample type.

The decision tree name (e.g., Postnatal in **Figure 351** above) will be displayed as the column name and values will be the classifications. The **Explanation** columns provide the reason or rule(s) that assigned the indicated classification. Checking off boxes for regions will display as columns for all user accounts. In **Figure 352** the **DECIPHER Syndromes Region** was selected to be displayed. Then, the **DECIPHER Syndromes** column is displayed in the **Sample Review** window, shown in **Figure 353**.

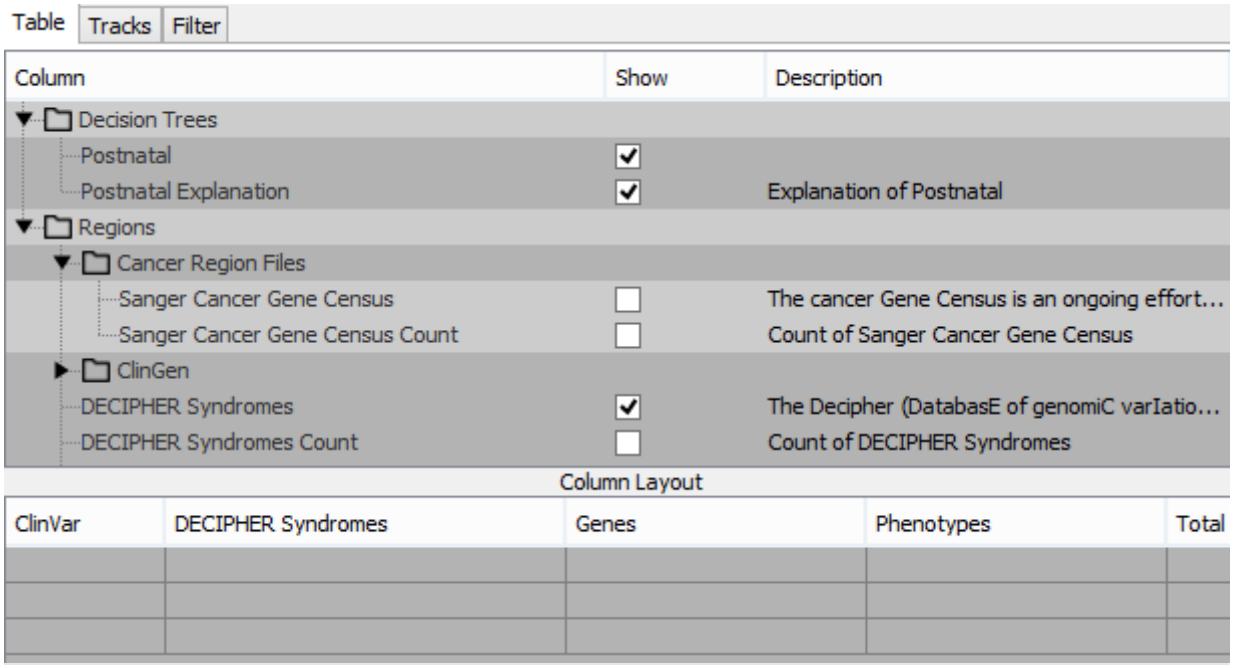


Figure 346. DECIPHER Syndromes selected.

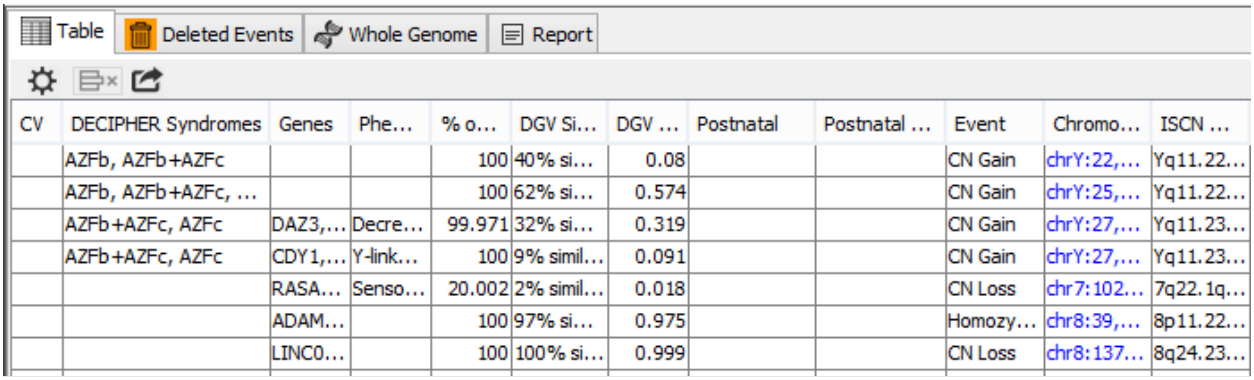


Figure 347. DECIPHER Syndromes column.

The order in which the columns will be displayed can also be set in the **Column Layout** pane by clicking on a column header and dragging the column left or right. After all selections have been made, clicking on the **Save Changes** button on the bottom right sets this as the default table display for all users.

The Tracks tab: Checking off the appropriate boxes will display Tracks for each sample type, seen in **Figure 354**. Set the track display order by clicking and dragging the track name up or down in the **Track Layout** section. Click on **Save Changes** on the bottom right.

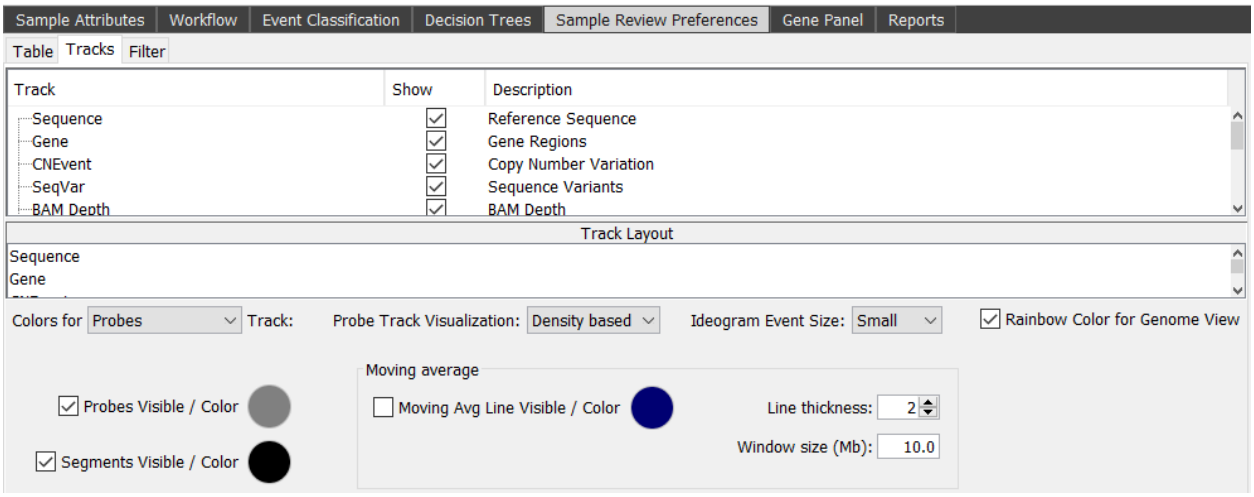
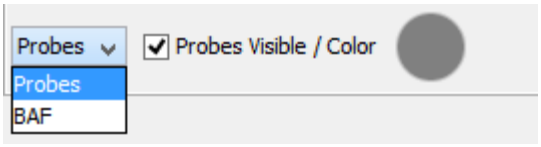


Figure 348. The **Tracks** tab.

At the bottom of the screen are selection boxes to hide/display certain plots and lines and the option to change the display colors. Click the dropdown to select for **Probes** or **BAF plots** and check off the boxes to display these plots:



To change the display color, click on the appropriate colored dot to bring up the color chooser and make a new selection, shown in **Figure 355**.

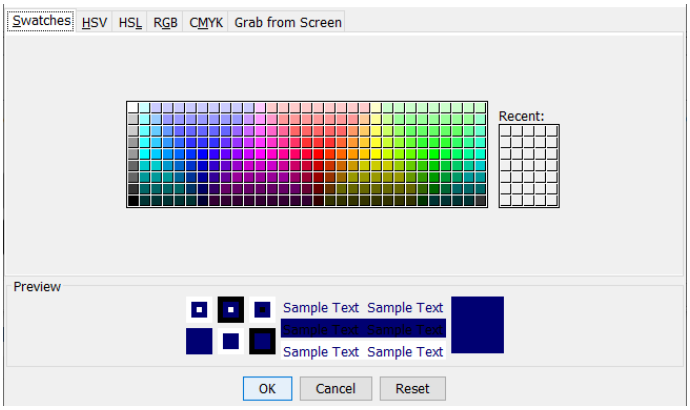


Figure 349. Color chooser.

There are tabs for the different color systems to specify a new color or one can also select a color from the screen via the **Grab from Screen** tab, seen in **Figure 356**. Simply click on the **magnifying glass** and drag over a color on the screen and release the mouse button. The new color will be displayed in the square while the current color is displayed in the circle. Click **OK** to select that color. The color preferences for the tracks and the probes plot after changing the probes color to purple is an example seen in **Figure 357**.

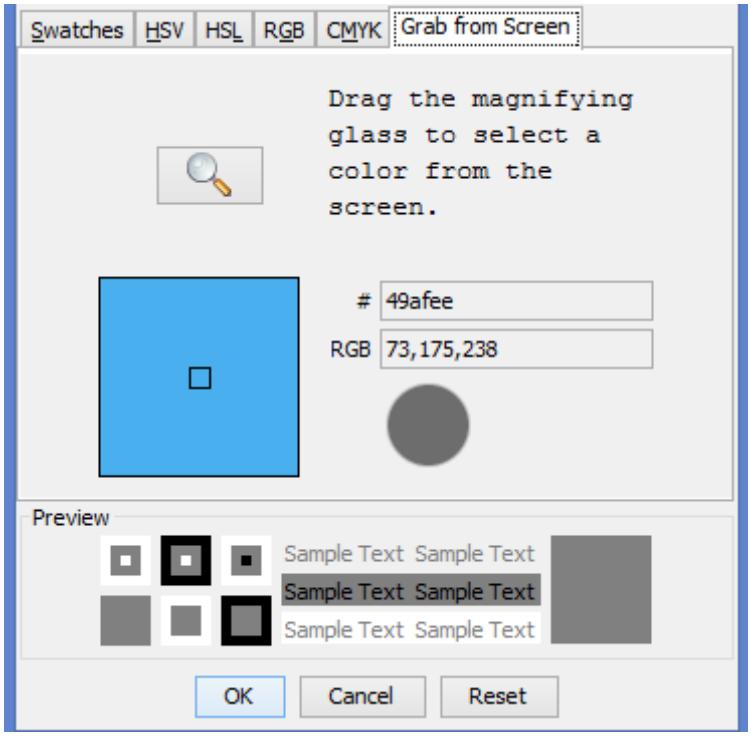


Figure 350. Grab from Screen tab.

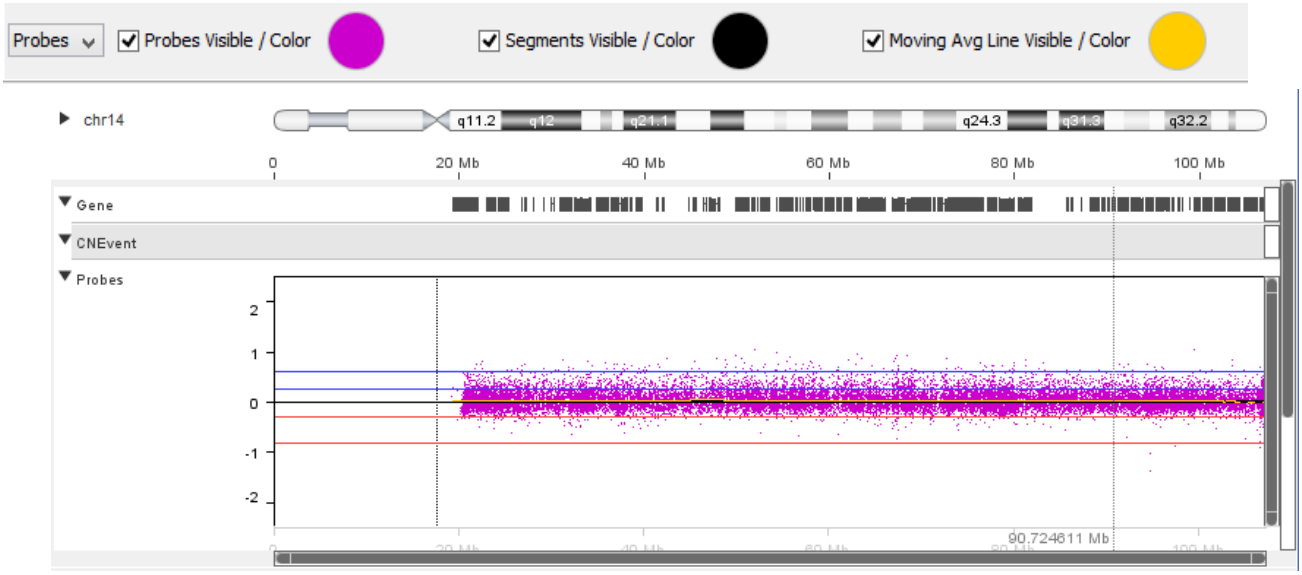


Figure 351. Purple probe color.

The Filter tab: Used to set default filters for the different sample types, various types of events can be removed or selected for display by marking off the appropriate checkboxes. Settings here are the default values set by the Admin, but each user can change the display in the **Sample Review** window.

Filtering is sequential following the chain displayed in the pane. The Admin can enable certain filters and make selections for the parameters to be applied to all samples of a specific sample type. Refer to the section on *Filtering of CNV, Allelic Events, and Sequence Variant Data* for detailed guidance on use of the filters.

The filters are grouped into four tabs, three representing each modality (**Copy Number**, **Allelic**, **Sequence Variants**, **Structural Variants**) and one for miscellaneous settings (**Other Settings**). Clicking on the **Gear** icon (indicated below with a red circle) in a filter box will display the available settings in the pane on the right. **Figure 358** shows the copy number filter chain with settings for classified events in the right panel.

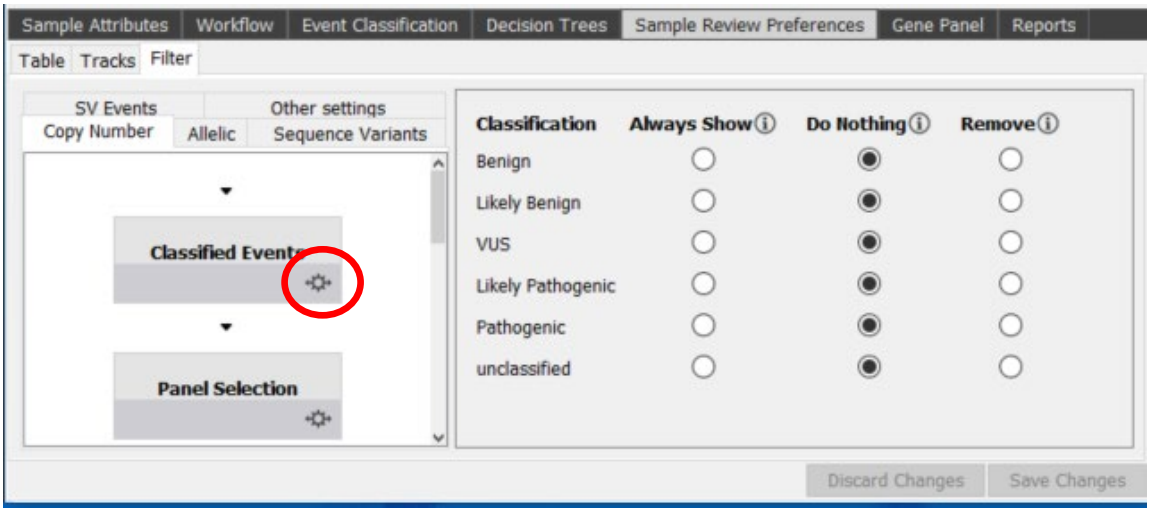


Figure 352. CN filter chain.

Make the appropriate selections for each filter and click **Save Changes** to apply them to all samples of the sample type being configured.

Figure 359 shows the **Sequence Variants** filter chain with settings for **Event Consequences Parameters** in the right panel.

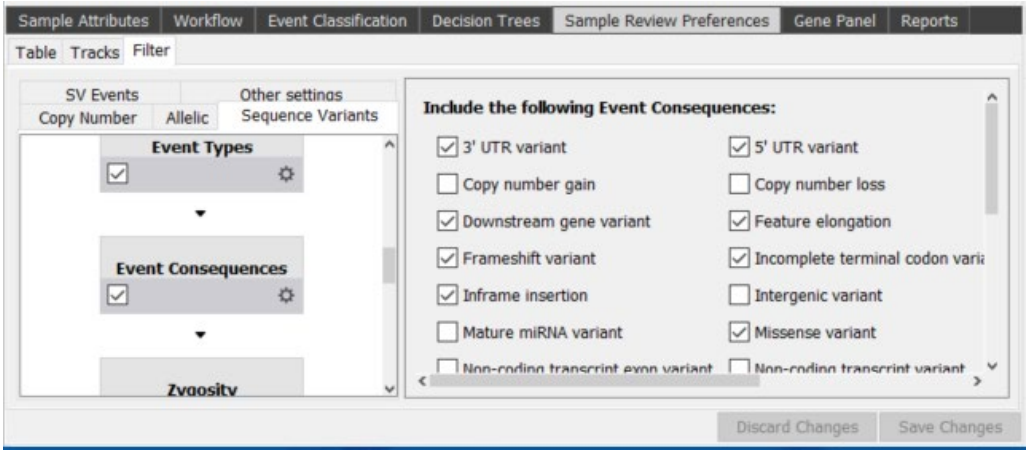


Figure 353. Event Consequences.

The **Panel Selection** locking option can be displayed by clicking on the **Gear** icon for, seen in **Figure 360**. Selecting a panel in the dropdown will default to the one used during sample review.

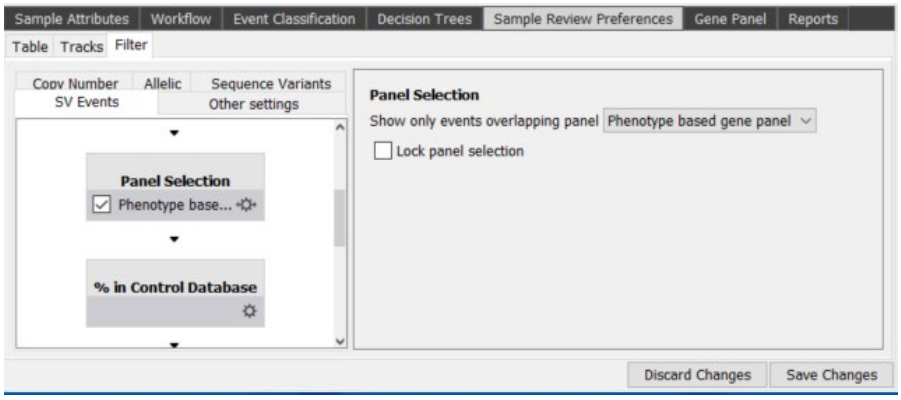


Figure 354: Panel Selection Filter image.

Clicking on will only display the selected panel to non-admin users. Users will only be able to select this panel for filtering. They will also not be able to upload ad-hoc panels or apply the phenotype based panel. Once the Admin selects a panel from the dropdown and marks off the **Lock Panel Selection** box, the **Panel Selection** checkbox will be marked and grayed out indicating that no changes can be made, as seen in **Figure 361**.

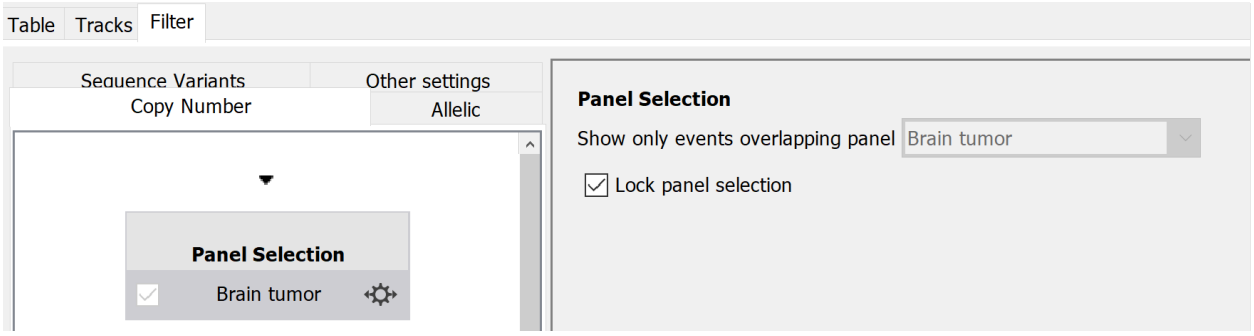


Figure 355. Lock panel selection.

The **Other Settings** tab includes additional filters for the Administrator to configure. The **Previous Cases Query** parameter, shown in **Figure 362**, is not part of the chain in the filtering UI but can be accessed in the tool panel of the **Browser** pane in the **Sample Review** window.

The Administrator can select the default filters for the sample type. However, during review in the **Sample Review** window, the user can make changes for an individual sample or for all samples of a particular sample type.

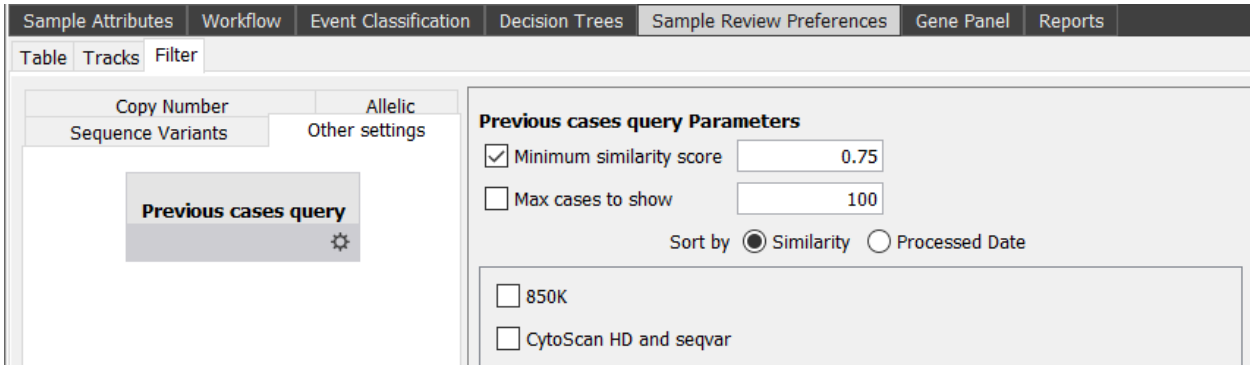


Figure 356. Parameter not included.

The Variant Details tab: Here the Admin selects which tracks will be used in the Variant Details Region overlaps to provide information on Genes and Regions by marking tracks as **Pathogenic**, **Benign**, or **None**. Tracks marked as None are not displayed in the **Variant Details** UI. Annotations displayed in **Variant Details** are tied to the test type.

Test Type = Constitutional has sections:

- Pathogenic Region Overlaps
- Benign Region Overlaps
- Gene Details

Test Type = Oncology has sections:

- Benign Region Overlaps
- Gene Details

Before the Admin marks tracks as Pathogenic or Benign, a set of defaults is assigned to all the tracks. The following logic is used internally in the software to set the defaults:

- Region tracks marked as Pathogenic, Likely Pathogenic, and VUS will be displayed in the **Pathogenic Region Overlaps** section.
- Those marked as Benign or likely Benign will be displayed in the **Benign Region Overlaps** section.
- Gene tracks marked will be displayed in the **Gene Details** section.

As an example, **Figure 363** shows some tracks with selections. These are non-cancer tracks so will be associated with Test Type=Constitutional.

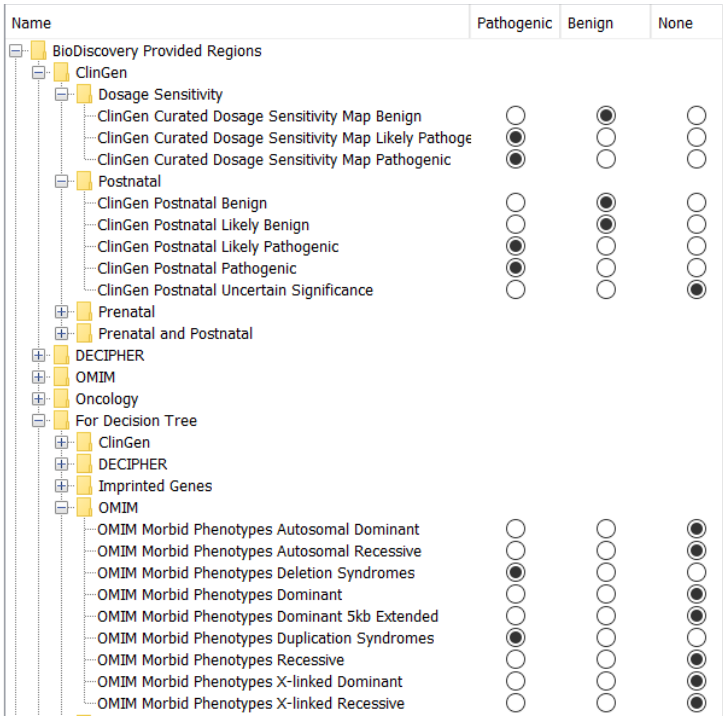


Figure 357. Test Type=Constitutional.

This corresponds to the fields in the **Variant Details** tab, shown in **Figure 364**.

NOTE: if there is no content for a specific variant in a track, the track will not be displayed.

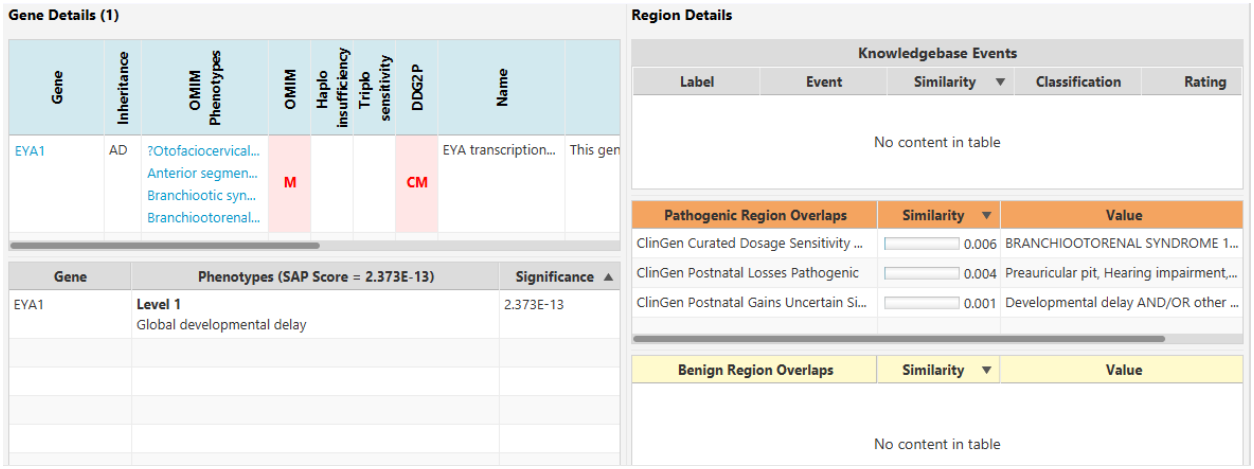


Figure 358. Corresponding **Variant Details** tab.

A sample with Test Type = Oncology will not display the Pathogenic Region Overlaps but will have CIViC in the **Gene Details** section.

ADMIN CREATION OF GENE PANELS

Importing genes/regions panels with transcript minimum read depth: The panel can be used to determine BAM coverage by specifiying the minimum read depth for genes/regions. The panel file must contain tab delimited columns with the following column headers:

Gene/Region Transcript Min Read Depth

If the header doesn't match the keywords above, the file will not be loaded and an error message will be displayed. The software recognizes that the import list also contains transcripts and read depth based on the presence of the column header **Gene/Region**. If there is a typographical error in this column header, the software assumes only genes/regions (no transcript or min read depth information) are being loaded and will attempt to load all values (concatenated per row) as genes/regions only.

If a row has a value for only one of the three columns, the software interprets this as the **Gene/Region** value even if the value is in one of the other columns. **Figure 365** is an example file containing transcripts and Read Depth.

Gene/Region	Transcript	Min Read Depth
BRAF	NM_001354609	20
CDKN2A	NM_058195,NM_000077	20
TP53	NM_001276697,NM_001276698	50

Figure 359. Example file.

NOTE: multiple transcript IDs must be separated by commas (recommended) or it can be one of the following as listed in the Secondary Delimiter field: **Tab**, **Comma**, **Semicolon**, **Pipe**, **Space**, **Tilde**. Make sure to select the correct secondary delimiter upon file import.

Once import is complete, the genes/regions will be listed in the right side-pane of the **Gene Panel** tab, as shown in **Figure 366**.

Gene/Region	Validated Region	Transcript	Min Read Depth
BRAF	chr7:140,419,126...	NM_004333	55
EGFR	chr7:55,086,713-...	NM_001346900	50
CDKN2A	chr9:21,967,750-...	NM_000077	
TP53	chr17:7,571,719-...		50
EGFR-AS1	chr7:55,247,442-...		300

Figure 360. Gene/Region.

If Min Read Depth is specified but a transcript is not, the software will take the entire gene (taking the union of all exons for the gene) for coverage calculation. This means that if there are no or very low coverage in the intronic regions, the coverage percentage will be extremely low.

TP53 – no transcript specified: no or very few reads in the intronic regions but good coverage on exons leads to a low coverage percentage (18.82%), as seen in **Figure 367**.

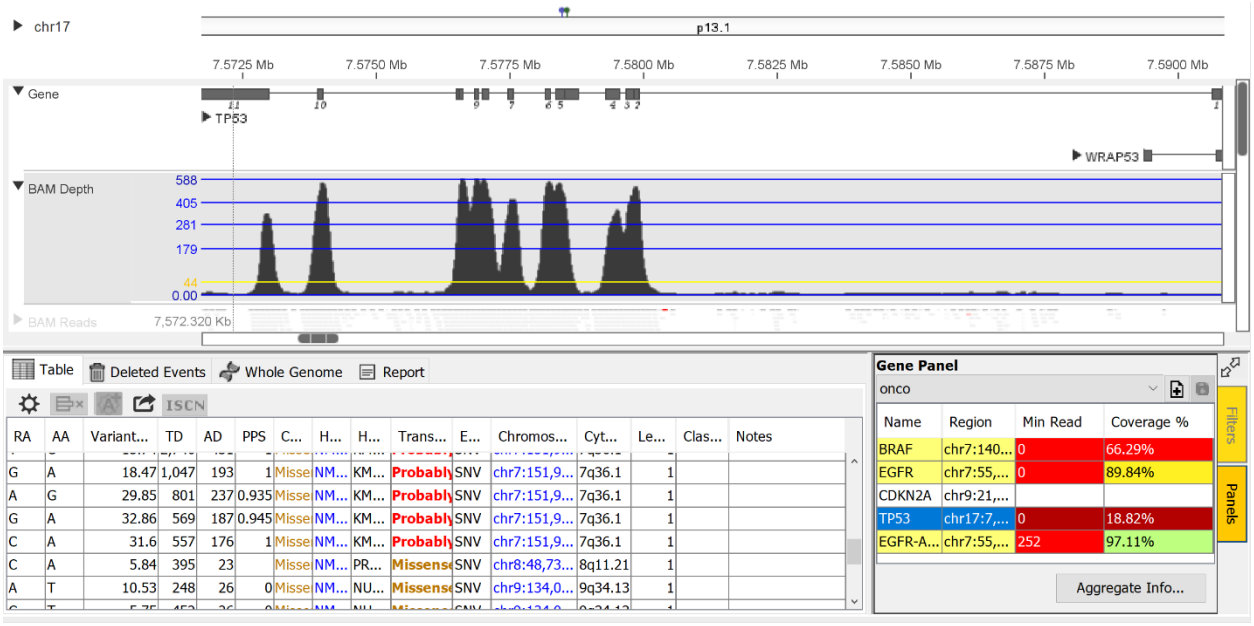


Figure 361. TP53.

If a transcript ID is specified without minimum read depth, then coverage will not be calculated for that gene/region (e.g., CDKN2A in the figure above).

Importing Exons: Another option for importing a list with specific regions and minimum coverage specifications is to import a list of exons per gene rather than transcripts per gene. The column headers are the same as when specifying transcripts but instead of listing transcript IDs in the Transcript column, exon regions can be specified in the format chr:start-end, for each exon as seen in **Figure 368**.

Gene/Region	Name	Transcript	Min Read Depth
APOE	APOE	chr19:45409122-45409184,chr19:45409855-45409928,	10

Figure 362. Importing exons.

Upon import, the software will validate the gene and list the gene region in the **Validated Region** column and the specified exons in the **Transcript** column:

Gene/Region	Validated Region	Transcript	Min Read Depth
BRAF	chr7:140,419,126-140,624,728	chr7:140432361-140434567, chr7:140439610-140439745, chr7:1404...	20
CDKN2A	chr9:21,967,750-21,995,042	NM_058195, NM_000077	20

Loading exons constricts coverage evaluation to only specified exons rather than all exons in a gene and can be important when only certain exons in genes are to be evaluated.

Importing a list of genes/regions in BED format: Click on the **Import** icon to bring up the file chooser. Select BED file in the **Files of Type** dropdown. This selection will display both .bed and .txt files since the BED format can be saved as a .txt file. As the BED format is unique, the software needs to know which parser to use so BED file must be selected for files saved in this format whether the extension is .bed or .txt. See **Figure 369**.

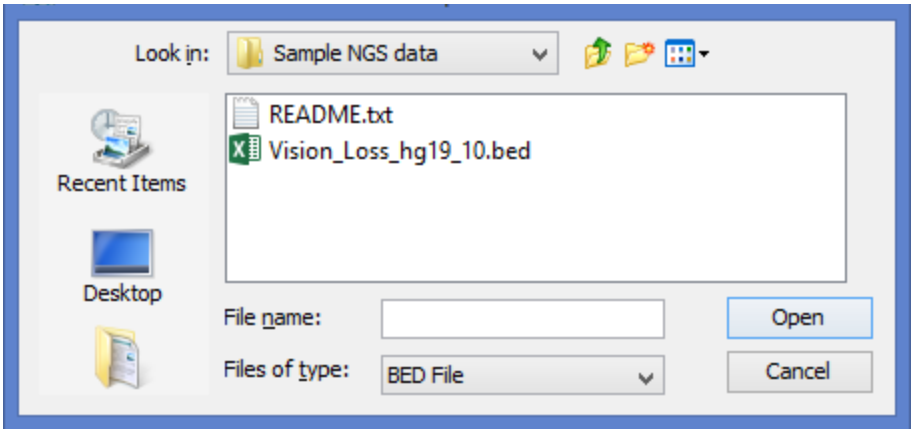


Figure 363. File chooser for BED format.

Select the file and click **Open**. The parser will validate the files and then display genes or regions in the right panel, as seen in **Figure 370**.

NOTE: BED files use the zero-based coordinate system so a start position of 0 in a BED file means that the region starts at base 1. The VIA browser is one-based and automatically corrects for this by adding 1 to the start position. Details on BED format are found here: https://genome.ucsc.edu/FAQ/FAQformat.html.

An example BED file with regions:

track db="hg19" name="Vison_Loss_10"			
chr19	3769092	3772219	RAX2
chr1	50574637	50667054	ELAVL4
chrX	153409744	153424505	OPN1LW
chr6	42123173	42147792	GUCA1A
chr4	120415550	120549981	PDE5A
chr3	139236283	139258490	RBP1
chr2	98962617	99015064	CNGA3
chrX	18657809	18690229	RS1
chrX	41306686	41334963	NYX
chr16	81272295	81324747	BCMO1

The BED file loaded into VIA as a gene panel:

Genes and regions
chr19:3,769,093-3,772,219
chr1:50,574,638-50,667,054
chrX:153,409,745-153,424,505
chr6:42,123,174-42,147,792
chr4:120,415,551-120,549,981
chr3:139,236,284-139,258,490
chr2:98,962,618-99,015,064
chrX:18,657,810-18,690,229
chrX:41,306,687-41,334,963
chr16:81,272,296-81,324,747

Figure 364. Genes or regions are displayed.

NOTE: A 1 was added to the start position to convert from the 0-based BED file to the 1-based coordinate system used in the VIA browser. The end positions are the same in 0-based and 1-based coordinate systems.

If an entered gene or region is not found by VIA, an alert will be displayed. If the entered gene is an alias, a message will indicate that the gene symbol entered will be replaced by the main gene symbol. For example, the user entered value is PARK2 and the software states that it is an alias and will be replaced by the official gene symbol PRKN, shown in **Figure 371**.

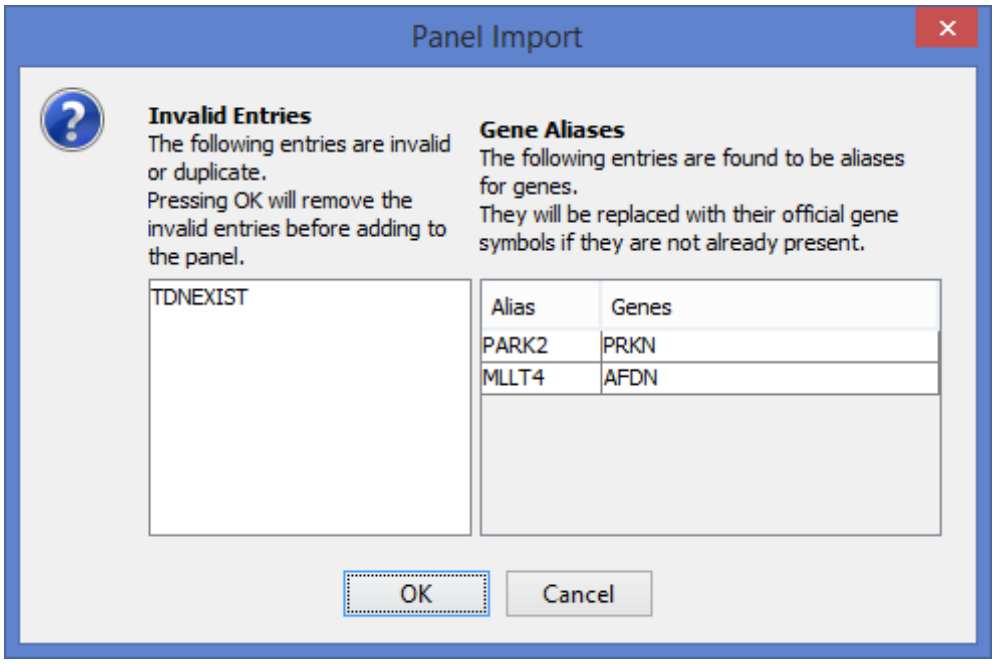


Figure 365. Alias

Validating a Legacy/Older Gene Panel: Version 5.0 introduced validation of gene panel regions which are then saved with the panel, but this validation was not done for gene panels added prior to version 5.0. For such panels, when the Admin first logs into VIA, a pop-up window will be displayed, seen in **Figure 372**, showing panels which have not been validated and providing the Admin an option to validate at that time or ask again later.

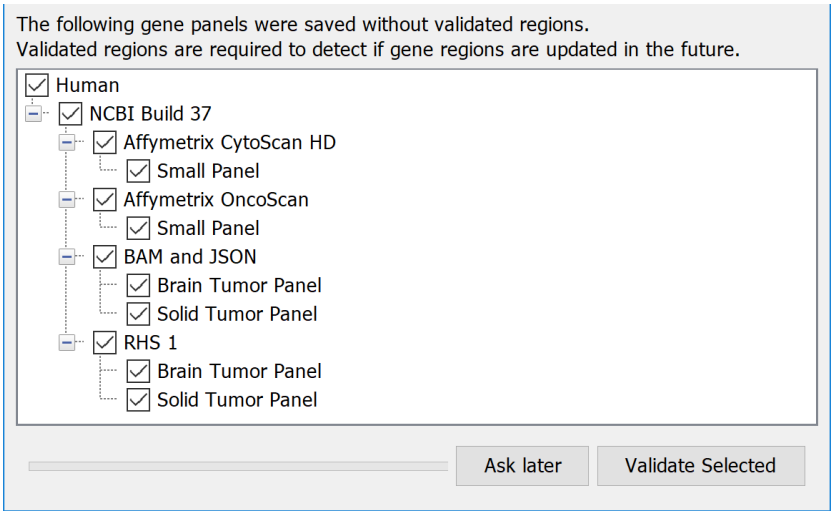


Figure 366. Validation.

Updating a Gene Panel with Changed Locations: If gene locations have changed (e.g., due to an annotation update), the **Gene** Panel must be updated to reflect the new regions. The user will receive an alert when locations have changed and will have to ask the Admin to update the panel. **Figure 373** is the message displayed when reviewing a case.

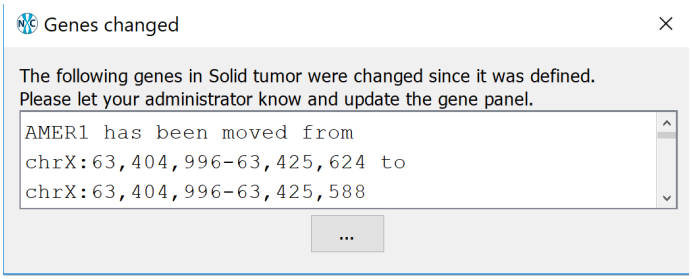


Figure 367. Changed locations.

The genes/regions can be updated manually one by one by editing the region. Genes can be removed and then re-added one at a time, or if the import file contained a list of gene symbols, that file can be re-imported to replace the existing list of genes; the regions will be validated and the new updated locations will be used.

Technical Reports (PDF): Different technical report templates can be set up for each sample type using the **Reports** tab to produce a PDF of specified results. First select the sample type at the top and then click the **+** button under the **Reports** tab to create a new report template. Enter a name for the report in the resulting window. Mark off the appropriate checkboxes to specify information to include in the report, as illustrated in **Figure 374**.

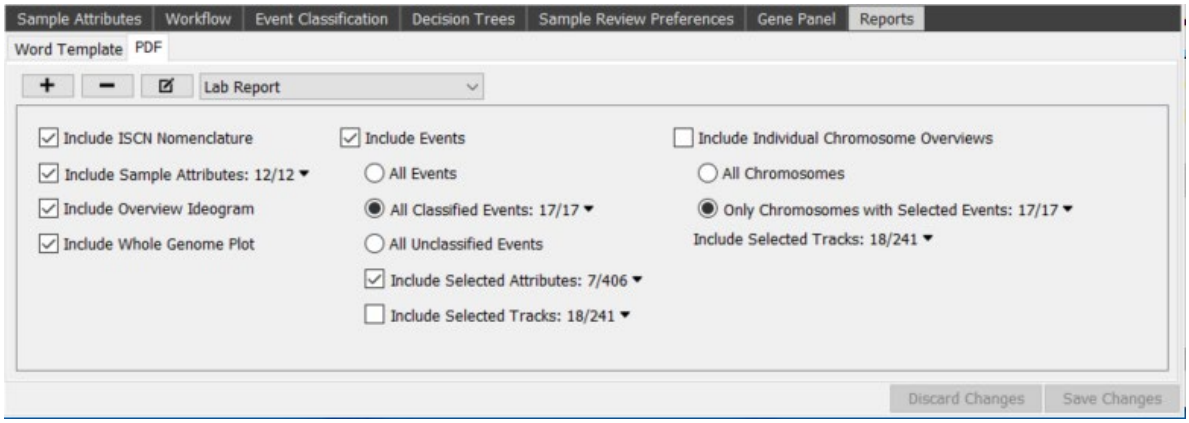


Figure 368. Selectable information for PDF reports.

Clicking on the black menu arrows displays selection boxes; check the boxes to display the items in the report. The numbers to the right of the black arrows indicate how many items are checked out of the total available.

When all selections are complete, click **Save Changes** on the bottom right to save the report template. If multiple report templates are specified for a sample type, the user can select which template to use using the **Report** tool when reviewing a sample. All available report templates will be listed in the **Report** tool dropdown.

NOTE: PDF reports are not supported for **Structural Variant** events.

Word Template Reports: VIA will populate a Word Template (.DOCX) file with event and sample level information, as seen in **Figure 375**.

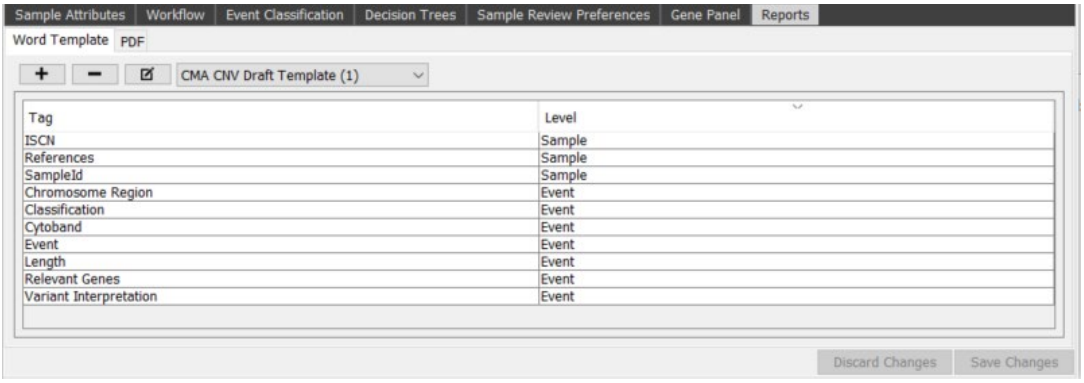


Figure 369. Example tags associated with Word template reports.

Tags: different components of a report template where tags may be added include main body, header/footer, repeating paragraphs, and tables. Tags can be added at the sample level or event level. Virtually any field/attribute associated with a sample or event can be added as a tag.

Sample Level Tags: These are based on the attributes associated with a sample type, including custom attributes added by the Admin and are denoted by **S:**. Note that the field **Sample** is displayed in the software UI.

Examples: «S:SampleId» «S:Gender» «S:ISCN» «S:Phenotypes»

Event Level Tags: These are based on the event table columns in the **Single Sample Review** window and are denoted by **E:**. Any existing column (as displayed in the **Table Preferences** window) can be used.

Examples: «E:Event» «E:Classification» «E:Genes» «E:Variant Interpretation»

Tags in Tables: Tags added in a table are populated with rows repeating for each event in the sample. The layout specified in the table in the template is repeated for each event. For example, **Table 21** from a Word report template demonstrates these tags. The final report would result in the populated **Table 22**.

Table 21. Tags in Tables.

Classification	Region	Length	Cytoband	Gene Count	Parent of Origin
«E:Classification»	«E:Chromosome Region»	«E:Length»	«E:Parent of Origin»	«E:Gene Count»	«E:Genes»
Genes: «E:SAP Score»					

Table 22. The resulting report.

Classification	Region	Length	Event	Cytoband	Gene Count
VUS	chr17:37,958,221-38,222,220	264000	Gain	17q12	7
Genes: TBC1D3L, TBC1D3D, TBC1D3C, LOC101929950, TBC1D3E, TBC1D3, NPEPPSP1					
VUS	chr7:159,060,487-159,345,973	285487	Gain	7q36.3	1
Genes: VIPR2					
Likely Pathogenic	chr22:19,724,872-19,917,026	192155	Loss	22q11.21	4
Genes: GNB1L, RTL10, TBX1, TXNRD2					

Tags in Repeating Paragraphs: Like the fields in a table that repeat, repeated paragraphs can be output. The content needs to be enclosed in repeat tags: «Repeat». The repeat tag must remain on its own line without any other text on that line, otherwise the surrounding text will be deleted when the paragraph is populated. Here is an example of a repeat paragraph in a template with the tags in bold.

RESULTS:

«Repeat»

A **«E:Event»** of **«E:Length»** bp on **«E:Cytoband»** was detected. This region overlaps the following genes **«E:Genes»**. This event has been classified as a **«E:Classification»** variant.

«Repeat»

When populated in the report, this paragraph section would be output as the following:

RESULTS:

A **Gain** of **264000** bp on **17q12** was detected. This region overlaps the following genes **TBC1D3L, TBC1D3D, TBC1D3C, LOC101929950, TBC1D3E, TBC1D3, NPEPPSP1**.
This event has been classified as a **VUS** variant.

A **Gain** of **285487** bp on **1=7q36.3** was detected. This region overlaps with the following genes **VIPR2**.
This event has been classified as a **VUS** variant.

A **Loss** of **192155** bp on **22q11.21** was detected. This region overlaps the following genes **GNB1L, RTL10, TBX1, TXNRD2**.
This event has been classified as a **Likely Pathogenic** variant.

When adding the tag for “Variant Interpretation” make sure to use this full term and not just “Interpretation”. “Interpretation” is used in the **Variant Details** tab whereas “Variant Interpretation” is used in the Table and both refer to the same field.

Special cases: For the OGM Heme Workflow, there are two additional tags that may be used. Both generate tables.

<**T:Guideline Variants**> inserts a table of the guideline targets that were imported for the sample type and columns for **Detected** and **Not Detected**. Example output:

Section A: Tier 1A Guideline Driven Variant Analysis for Acute Myeloid Leukemia Results

Variant	Detected	Not Detected	Variant	Detected	Not Detected
Chr5 whole chromosome (monosomy)	X		t(3;5) *NPM1::MLF1* (translocation)		X
5q (deletion; includes 5q31.2)		X	t(3;3) *GATA2::MECOM* (translocation)	X	
Chr7 whole chromosome (monosomy)		X	t(6;9) *DEK::NUP214* (translocation)	X	
7q (deletion; includes 7q31.2)		X	t(8;21) *RUNX1::RUNX1T1* (translocation)	X	
Chr11 *KMT2A*					

<**T:Whole Genome Results**> outputs a table with three columns. For example:

Template:

Chromosome	ISCN	Classification
«T:Whole Genome Results»		

Output:

Chromosome	ISCN	Classification
Chr 1	ins(1;?)(p36.22;?)(12054956_12054957;?)	N/A
	1p36.21(13039529_13040611)x1	N/A
	inv(1)(p36.21p36.21)(13040508_13216782)	N/A
	ins(1;?)(p36.21;?)(13080876_13080877;?)	N/A
	inv(1)(p36.13p36.13)(17028282_17185676)	N/A
	1p12p11.2(120574811_121266910)x1	N/A
Chr 2	None	
Chr 3	None	

How to Add Merge Fields

Many tutorials/instructions can be found online on how to add merged fields. Two examples are:

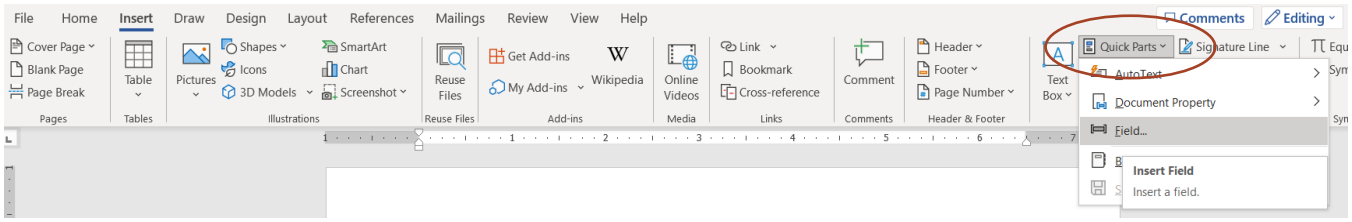
Merge fields for Windows

Merge fields for Mac

Merge fields for Windows

1. Open Microsoft Word and select the **Insert** tab.

2. Go to the **Quick Parts** button. Depending on the size of the Microsoft Word window, the button can be in its own column, or it might be stacked up with other buttons.
3. Click on the **Quick Parts** button.
4. Then select **Field.**



5. Select **MergeField** from the leftmost menu, as seen in **Figure 376**. **NOTE:** Selecting a different field type will make the template invalid.

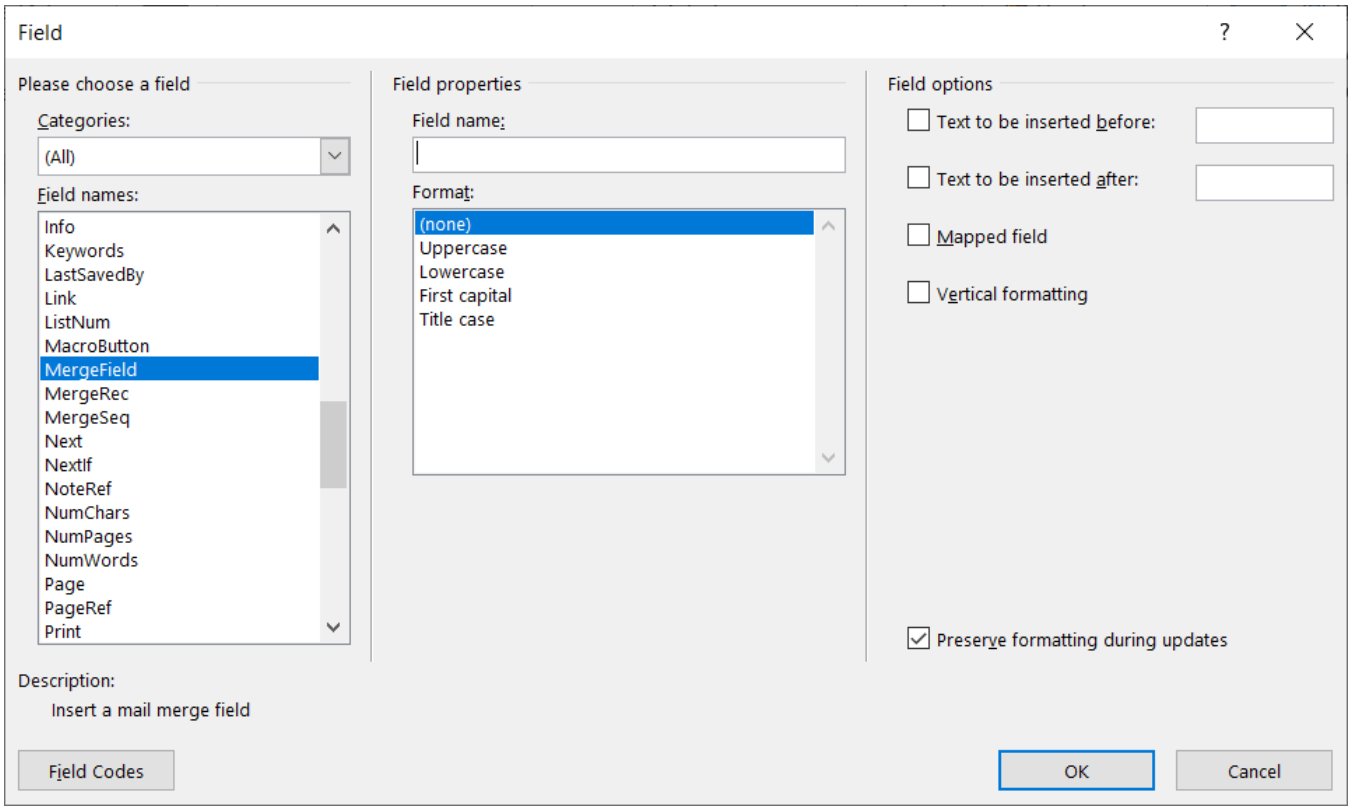


Figure 370. Field types.

6. Type in the name of the tag you want to insert, as seen in **Figure 377** (Example: S:SampleId, E:Event, etc.). Select **OK**.

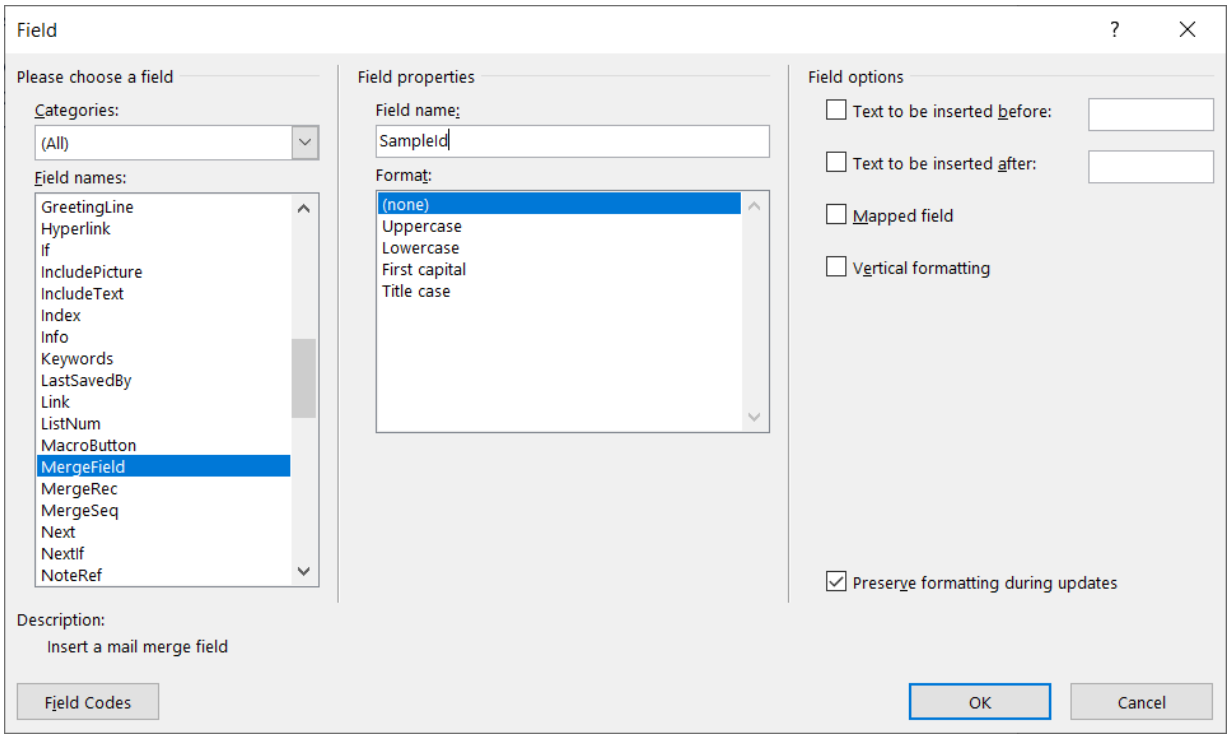
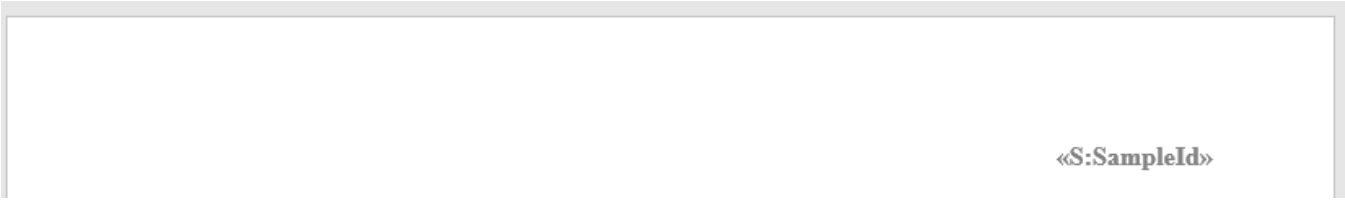


Figure 371. Field names.

The merge field has now been created. Insert white space (press **spacebar** or **Enter**) immediately following the inserted tag. Failure to add white space may cause formatting issues in the template, resulting in it becoming invalid.



Processing Usage: The administrator has access to a record of the processing jobs submitted to the processing unit. The list of processed samples is available within the **Processing Usage** tab, includes the date, duration, sample class, and user. The list of processed samples is exportable. Processed samples will be labeled as New, Reprocessed, or Duplicate. Only samples counted as New consume a sample credit for the associated sample class.

Sample credits in VIA are consumed in the following scenarios:

1) When a user processes a new sample.
2) When a user deletes an existing sample that has been previously processed, and re-uploads and processes the previously deleted sample as a new sample.
3) When a user overrides an existing sample and processes the sample.

Examples where a sample credit is not consumed:

1) Reprocessing an existing processed sample.
2) Duplicating a processed sample and processing the duplicate copy.

Technical Assistance

For technical assistance, contact Bionano Genomics Technical Support.

You can retrieve documentation on Bionano products, SDS's, certificates of analysis, frequently asked questions, and other related documents from the Support website or by request through e-mail and telephone.

TYPE	CONTACT
Email	support@bionano.com
Phone	Hours of Operation: Monday through Friday, 9:00 a.m. to 5:00 p.m., PST US: +1 (858) 888-7663
Website	www.bionano.com/support
Address	Bionano Genomics, Inc. 9540 Towne Centre Drive, Suite 100 San Diego, CA 92121