VariantGrid

CCB ACRF Cancer Genomics Facility

Oct 04, 2023

OVERVIEW:

VariantGrid is a source available variant database and web application for analyzing genetic data.

This documentation is intended for users. There are also Admin docs and Developers Technical Wiki

CHAPTER

ONE

INTRO

VariantGrid has a number of installations. Please visit the individual sites for login/registration details.

1.1 Cloud servers

- variantgrid.com - Research cloud server
- runx1db - Rare disease exome sharing
- Shariant - Australian Genomics variant classification sharing platform

1.2 Private server

There is a VariantGrid private server inside SA Pathology, the public pathology provider to the South Australian Health.

The advantages of a private server are being restricted to a private intranet, and being able to analyse private patient data without worrying about it being on the cloud.

To install a local copy of VariantGrid, please see the GitHub page.

CHAPTER

TWO

TECHNICAL ATTRIBUTIONS

Shariant, RunX1, SA Pathology VariantGrid and variantgrid.com are built upon VariantGrid technology.

2.1 Genetic/Medical Databases

Sources used for VariantGrid annotations:

- Ensembl Variant Effect Predictor (VEP)
- 1000 Genomes
- cadd
- ClinVar
- cosmic
- dbsnp
- Ensembl
- exac
- Exome Sequencing Project (ESP)
- fathmm
- gerp
- Human Protein Atlas
- Human Phenotype Ontology
- pfam
- phylop
- UCSC

2.1.1 Literature References

Sources used for summaries of cited literature:

- NCBI Bookshelf
- PubMed
- PubMed Central

2.1.2 Technical architecture

VariantGrid is open source, written in Python 3, and depends on many libraries. The main components are:

- Django
- Postgres
- RabbitMQ
- Redis
- Nginx
- Guinicorn

2.1.3 General Icons By

Icons made by Freepik from www.flaticon.comIcons made by Dave Gandy from www.flaticon.comIcons made by Google from www.flaticon.comIcons made by Smashicons from www.flaticon.comIcons made by Chanut from www.flaticon.comIcons made by Those Icons from www.flaticon.comIcons made by Good Ware from www.flaticon.comIcons made by mynamepong from www.flaticon.comIcons made by Darius Dan from www.flaticon.comIcons made by Linector from www.flaticon.com

2.1.4 Space Avatar Icons By

Icons made by Freepik from www.flaticon.comIcons made by Dave Gandy from www.flaticon.comIcons made by Smashicons from www.flaticon.comIcons made by Creaticca Creative Agency from www.flaticon.comIcons made by Pixel perfect from www.flaticon.comIcons made by Eucalyp from www.flaticon.comIcons made by Kiranshastry from www.flaticon.comIcons made by itim2101 from www.flaticon.comIcons made by Payungkead from www.flaticon.com

CHAPTER

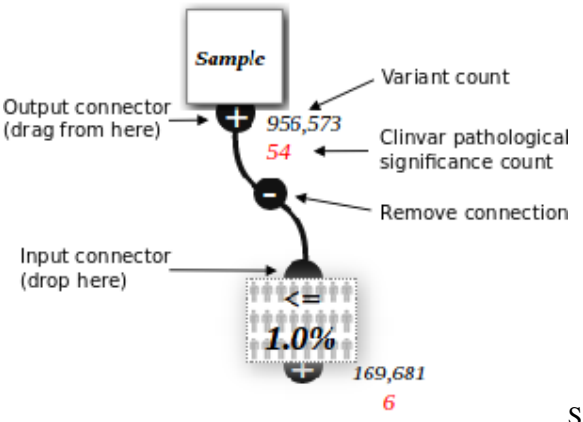
THREE

ANALYSIS INTRO

Create custom variant filters by connecting together nodes representing sources or filters of variants. See analysis nodes

Other variant databases allow similar creation of filters, but VariantGrid can constuct nodes in real-time, enabling rapid exploration of large and difficult genomic data sets.

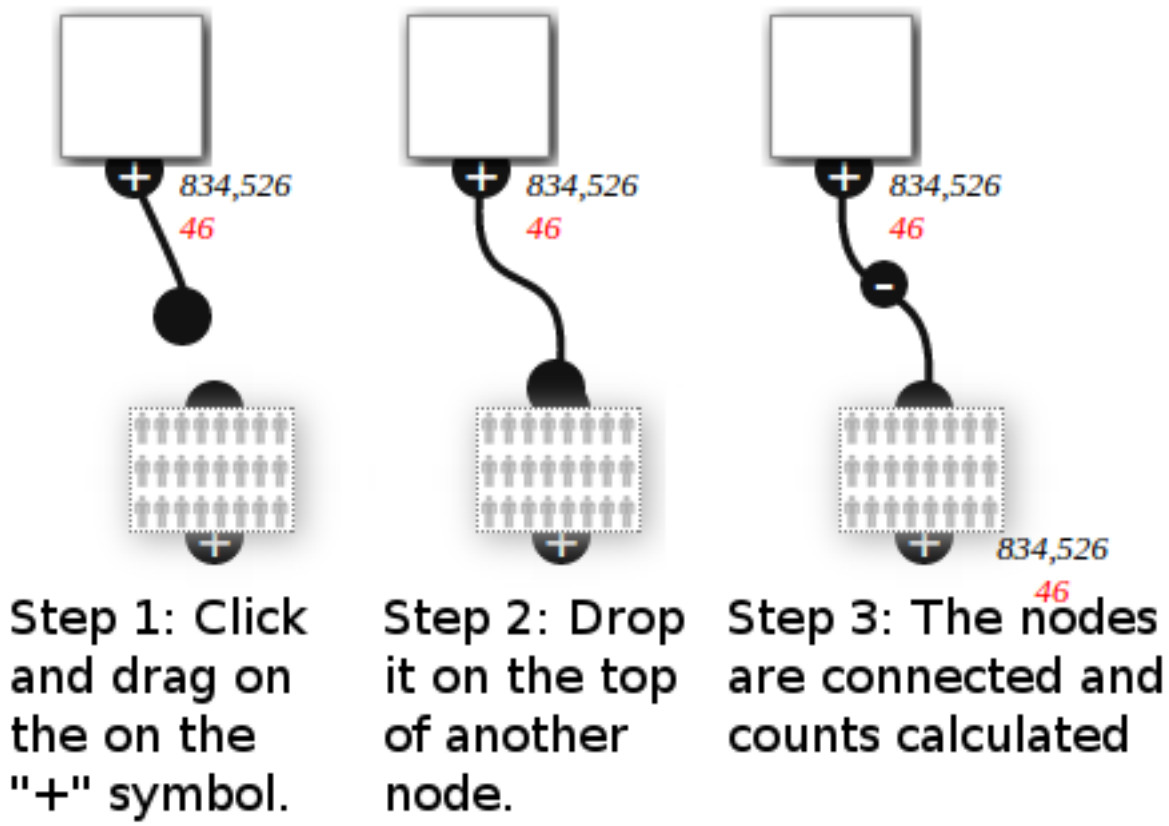
3.1 Analysis Nodes



Sample Node connected to a Population Filter Node

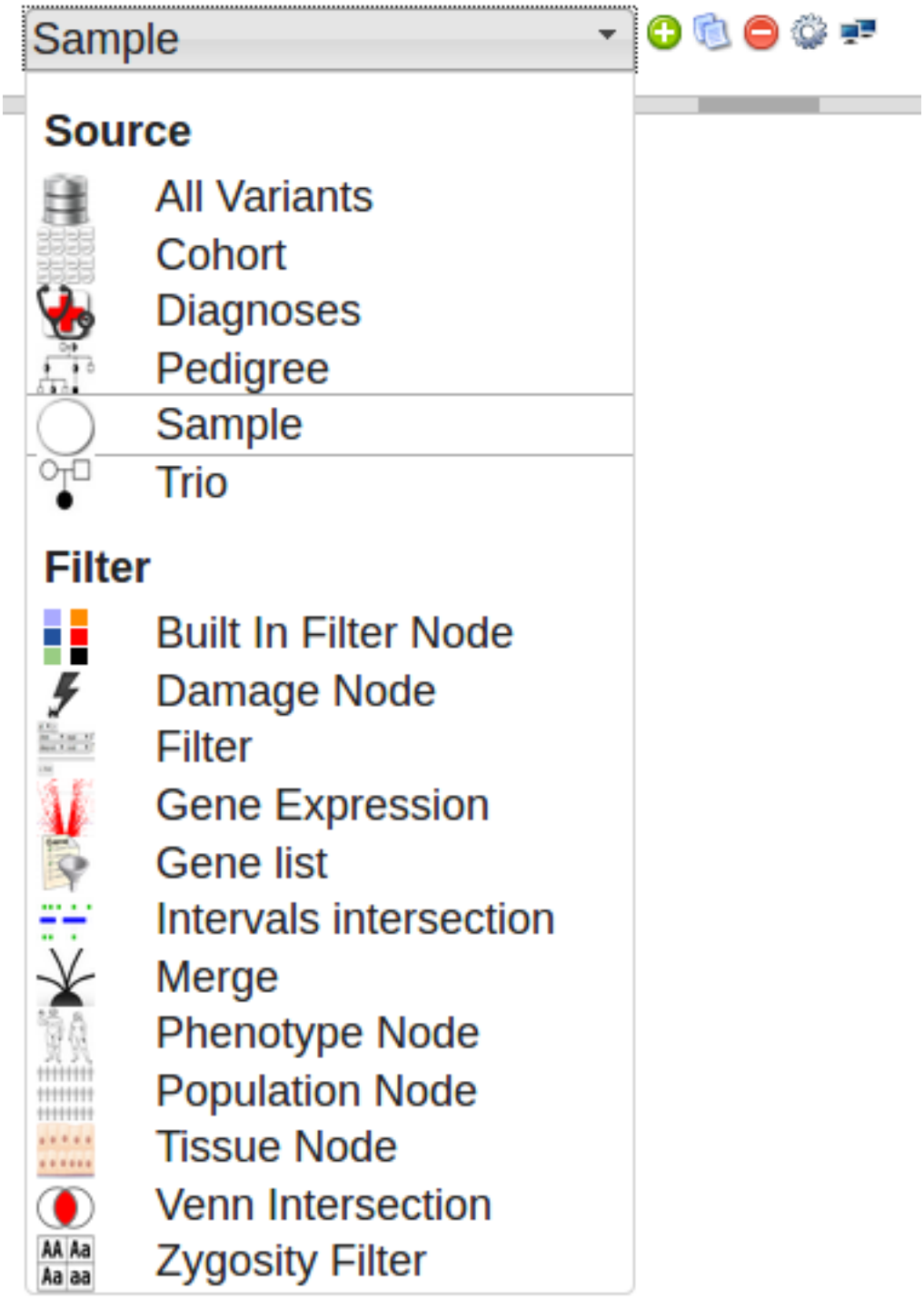
The top node is configured to show a particular patient exome (from an uploaded VCF).

These variants are then filtered to those that are less than 1% of the population.



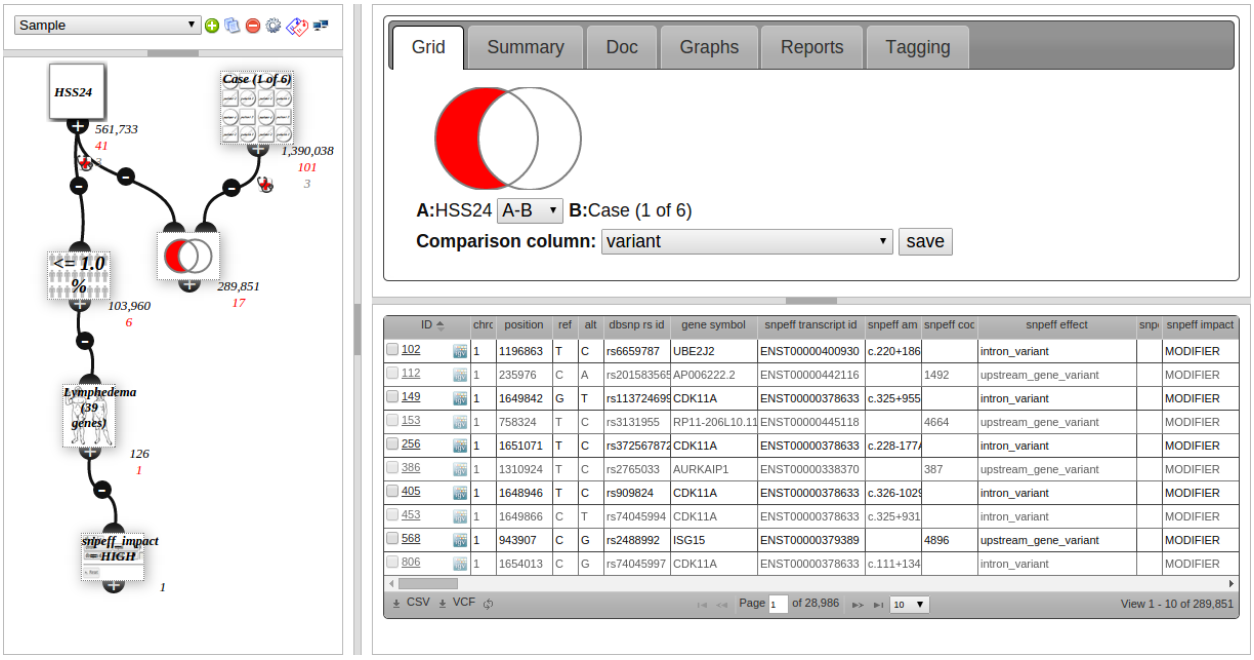
Connecting Nodes

To add a node, select the node type from the drop down menu in the top left of the screen and click the add button



Click and drag a node to move it around. You can select multiple nodes by drag-selecting a box around them. This allows you to copy, delete or move them as a group. Delete selected nodes by pressing DELETE, or click the delete button.

3.2 Analysis screen



The screenshot above shows the VariantGrid analysis screen. The node graph is on the left part of the screen, showing the user built filters.

Click a node to select it. This loads the node editor (top right) and a grid of the variants (see section below) in the node (bottom right).

Clicking on the node loads this editor window. The node editor is different depending on the type of node.

3.3 Analysis Grid

The 1st column (ID) is special and contains a check box, a numbered link and an IGV logo. The check box is used to select rows manually. The link loads detailed information about that variant above the grid. The IGV link will view the locus in IGV (loading bam files associated with samples). See IGV Integration page. Clicking on a row highlights it. Select the "tagging" tab, then click on a label to tag/colour the row.

CHAPTER

FOUR

ANALYSIS NODES

4.1 Source Nodes

Source nodes allow you to add variants to your analysis either by adding samples, groups of samples, or groups of variants from within the VariantGrid database. Each source node provides options to filter the variants available. Before changing the default filters available on the source nodes it's important to be familiar in interpreting variant zygosity and parameters (AD,DP,GQ,PL, AF) as these filters will have a marked impact on the variants displayed for analysis.

The following sections provide details of each of the different source nodes and associated filters available to curators.

4.1.1 All Variants

Retrieves all variants in the database. This can be restricted to a gene, or by zygosity.

Default is to show variants with a minimum of 1 of "any zygosity" (ie HET/HOM ALT) as this removes variants with unknown zygosity or variants that are not associated with samples in the database (eg from ClinVar)

To see all variants - "any zygosity" min to 0, but be aware that this will dramatically increase the results returned. Reference variants come from HOM_REF calls matching sample HET calls, low frequency somatic calls or multi-sample germline VCFs.

The node returns variants at the time it was saved (this "Last saved" date in the editor). Variants are constantly added to the system, clicking save may return more results than last time.

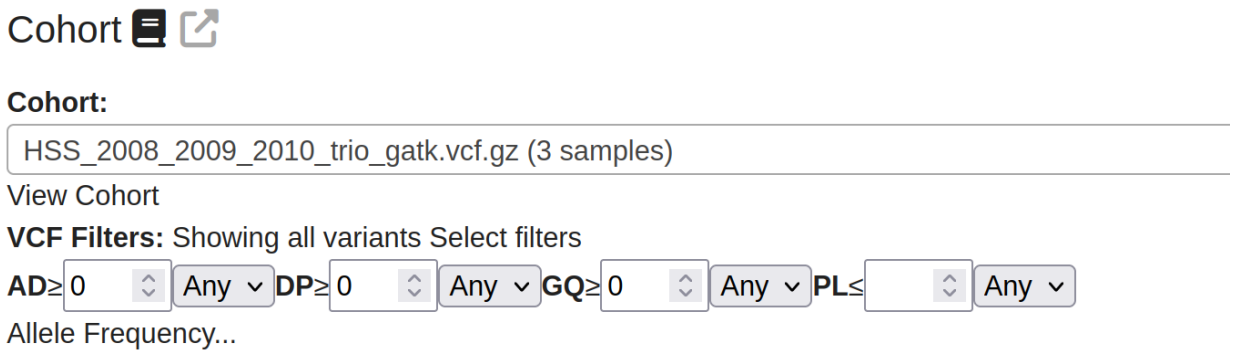
4.1.2 Cohort

Used to add a collection of related samples, eg “control group” or “poor responders”.

VariantGrid will automatically generate a cohort for each vcf upon upload. This cohort will contain all samples in the vcf. All other cohorts need to be defined manually by the user. Once defined, a cohort will be available for selection in the dropdown menu on the cohort node. It is recommended, though not essential, that samples to be analysed as a cohort are joint-called in the same vcf where possible.

There are two main approaches available to filter variants within a cohort:

Parameter Filtering: Filtering based on any combination of the variant parameters AD,DP,GQ,PL or AF.



After each parameter is All/Any - this sets whether the parameter must be at least 1 sample or all of them.

Note that not all vcfs will contain values for these parameters. Missing values will result in variants being inadvertently filtered from the cohort, so check your samples carefully before applying these filters.

Zygosity filtering: There are 3 methods for filtering cohorts by zygosity: zygosity counts, simple zygosity or sample zygosity. The selected method is the method that is expanded after the node filters have been saved.

Parameter and zygosity filtering can be applied together, however, only one zygosity filter type (count, simple or sample) can be applied at any one time. By default cohorts are filtered using only the simple zygosity method: Het or Hom_Alt for ALL samples.

Zygosity Counts

"Any Zygosity" = Hom/Het/Ref (ie anything other than 'unknown'). Unknown zygosity is when there is no coverage over the variant for this sample.

These counts are applied together in an AND-like manner. Warning: It's possible to set ref/het/hom alt minimums that add up to more than the number of samples in the cohort, which will always be false, and so exclude all variants.

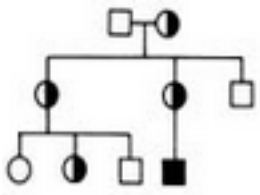
4.1.3 Classifications

The Classifications node is used to add internally classified variants to the analysis workflow. Use the checkboxes to display variants with classifications matching the selected clinical significance.

The 'other' checkbox includes the following: artefacts, drug response or risk factor.

If a variant has been classified multiple times with differing clinical significance it will be shown if any of the classifications match the selected clinical significance. For example, let's say the ASLX1 variant X has been classified as both an artefact and likely pathogenic (this situation may occur if a truly pathogenic variant can't be reliably sequenced on a specific platform, e.g. amplicon v capture). In this case Variant X will be displayed if either of the artefact or likely pathogenic tickboxes are selected.

4.1.4 Pedigree



Variants from a *Pedigree*, filtered by genotype according to Autosomal Recessive and Autosomal Dominant inheritance models.

Autosomal Recessive: Affected=HOM_ALT, Unaffected=HET **Autosomal Dominant:** Affected=HET or HOM_ALT

4.1.5 Sample

This node will load all variants present in a sample (equivalent to a single column in a vcf). A sample is usually one genotype (patient, cell or organism) with a set of variants.

This node is particularly useful for singleton analyses. Similar to the cohort node, a sample node can be filtered by variant parameters AD,DP,GQ,PL or AF (if available in the vcf), and also the variant zygosity. Before filtering by variant parameters make sure that they have been provided in the vcf otherwise no variants will be shown!

4.1.6 Trio

This node adds all variants present in a trio of samples. Trios need to be defined manually by the user. This includes specifing parental and proband samples, along with the affected status of the samples. Once defined, a trio will be available for selection in the dropdown menu on the trio node in the analysis workspace. It is recommended, though not essential, that samples to be analysed as a trio are joint-called in the same vcf where possible otherwise it is not possible to determine whether missing data is due to a reference call or a lack of coverage at the locus.

Each trio node requires an inheritance mode to be selected. This selection will then filter the variants according to the zygosities as listed in the table below. Only one inheritance mode can be selected per trio node. To assess multiple different modes of inheritance add multiple trio nodes to the analysis workspace. Use the default trio analysis template to quickly construct a trio analysis.

If “require parent zygosity” is False - parent zygosities may be “Unknown”. Selecting this option will allow variants with low or no coverage in parental samples to pass the zygosity filters. Note that if the samples have not been joint-called this may also allow parental reference calls through due to missing data.

Below is the table is “require parent zygosity” is True:

In addition to the above modes of inheritance the trio node can be used to filter a sample to compound het variants. To do so add the trio node below an existing workflow for a sample and select the compound het mode of inheritance. This filter finds common genes with *both* “het from mother” and “het from father” and zygosity of (het from mother OR het from father) as per the table below.

Note that the placement of the compound het filter within a workflow is important. If the node input contains too many variants or artefacts, many false positive compound het calls will be shown in the trio c.het node. Conversely, if the filtering has been too stringent, real compound het variants will be excluded.

Compound HET

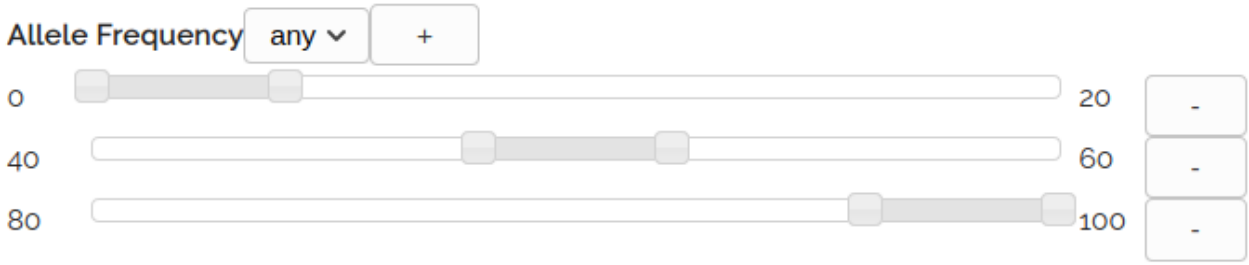
4.2 Filter Nodes

These nodes filter variants connected to the top of them

4.2.1 Allele Frequency

Filter based on a sample's variant allele frequency (AF). If multiple samples have been used in the analysis workflow, make sure to select the sample of interest using the dropdown in the node menu.

The AF is reported as provided by the vcf, if the AF is missing from the vcf VariantGrid will calculate the AF. Details on the source of the AF are provided in the vcf header, which can be viewed in the vcf info tab on the vcf details page (/snpdb/view_vcf/X)



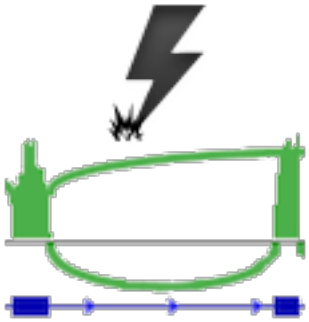
4.2.2 Built In Filter

The built in filter allows selection of commonly used variant classes including variants with:

- ClinVar - Variants with a ClinVar Max classification of Likely Pathogenic or Pathogenic
- OMIM Phenotype - Variants in genes with an OMIM phenotype
- HIGH or MODERATE IMPACT - Variants with a HIGH or MODERATE IMPACT as predicted by the VEP pipeline
- Classified - Variants that have been classified in VariantGrid with any clinical significance

- Classified Pathogenic - Variants that have been classified in VariantGrid with a maximum clinical significance of Likely Pathogenic or Pathogenic
- COSMIC - Variants reported in the COSMIC database (COSMIC count > 0)

4.2.3 Effect



The effect node allows for quick filtering of variants based on a combination of predictions and information sets.

To enable any of the pre-set filters, click the left checkbox then move the slider to select variants meeting or exceeding the set threshold (T). By default, if multiple filters are selected variants will be shown that meet **ANY** of the of the criteria. It is recommended to **ALWAYS** include IMPACT min = HIGH in a basic filter set as this will prevent inadvert loss of loss of function variants (frameshift/splice donor/start loss/stop gain etc.) that lack prediction data.

AVAILABLE FILTERS

Impact min Allow variants with an impact greater or equal to the selected impact level. Impact levels are ordered as follows: MODIFIER < LOW < MODERATE < HIGHFor example, impact min = LOW will display variants with IMPACT = LOW or MODERATE or HIGH

The MODERATE* filter is a special case developed to exclude missense variants. The MODERATE* filter was designed so that curators can quickly remove tolerated/benign missense variants. **It is recommended to always use the MODERATE* option in combination with one or more of the REVEL, CADD or Damage Predictor options to control which missense variants will be displayed.** Specifically MODERATE* will display variants as follows:

- Any variants with IMPACT = HIGH plus
- Any variants with IMPACT = MODERATE and VARIANT CLASS != SNV

As an example, test filtering your dataset using only the MODERATE option. You will see that all missense variants are displayed (along with MODERATE indels/substitutions and all HIGH impact variants). Many of the missense variants have low pathogenicity predictions and no other data to indicate they are deleterious. These variants are normally discarded by curators upon review. To speed up this process, now trying filtering your dataset using the MODERATE* option + REVEL min = 0.7. Now you will see that the only missense variants displayed are those with REVEL scores greater or equal to 0.7. These are your missense variants of interest. Because you've chosen the MODERATE* filter you'll still see indels/substitutions with MODERATE impact along with all HIGH impact variants.

Splice min Variants meeting the following criteria will be displayed:

- dbscSNV.ADA >= T or
- dbscSNV.RF >= T or
- SliceAI.DL.Score >= T or
- SpliceAI.DG.Score >= T or

- SpliceAI.AL.Score >= T or
- SpliceAI.AG.Score >= T or
- is splice indel

Where a splice indel is defined as: (splice region is not null AND variant class is not SNV). Splice indels have been included to ensure that insertions, deletions and complex variants in a splice region are not removed by the filter as these variants are not generally assessed by splicing predictors. As a rule of thumb a splice threshold of 0.2 is lenient, 0.4 moderate and 0.6 stringent.

For further information on these splicing predictors see:SpliceAI: https://pubmed.ncbi.nlm.nih.gov/30661751/dbscSNV: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4267638/

CADD score minCADD phred >= T

REVEL score minREVEL score >= T

COSMIC count minCOSMIC count >= T

Damage predictions minsum(pathogenic predictions for variant) >= T

A prediction is considered pathogenic if it meets the following criteria:

- SIFT = damaging
- Polyphen2 = possibly or probably damaging
- Mutation assessor = medium or high
- Mutation taster = disease causing
- Fathmm = damaging

Protein domain If selected, this will display variants with values in at least one of the following fields:

- Interpro_domains
- domains

PublishedIf selected, this will display variants with values in at least one of the following fields:

- Pubmed
- MM variant article count
- MM variant/protein article count
- MM aa article count
- MM AA ID

FILTERING EXAMPLES

Using the following 2 variants as an example:

Example 1: Filter Set: CADD 20; REVEL 0.7; IMPACT MODComputed as: CADD >= 20 OR REVEL >= 0.7 OR IMPACT >= MODResult: Both Variant 1 & 2 will displayed.

Advanced use of effect node filters: Click on the required link to display required and null checkbox options. Warning: do not use these checkboxes unless you are comfortable with Boolean logic and the behaviour of null data for your selected filters. If a criterion **MUST** be met to display a variant, select the required box for each required criterion. Make sure to check the **"Allow Null"** box if results should include variants with missing data for the selected criterion. It is particularly important to check the 'Allow null' box if REVEL or CADD scores are set to 'required' otherwise all indels will be filtered as predictions are only available for SNVs. Below are some advanced examples using the variants from the table above:

Example 2: Filter set: CADD 20; REVEL 0.7 (required); IMPACT MODComputed as: REVEL >= 0.7 AND (CADD >= 20 OR IMPACT >= MOD)Result: No variants will displayed.

Example 3: Filter set: CADD 20 (required, null); REVEL 0.7; IMPACT MODComputed as: (REVEL >= 0.7 OR REVEL is null) AND (CADD >= 20 OR IMPACT >= MOD)Result: Only variant 2 will be displayed.

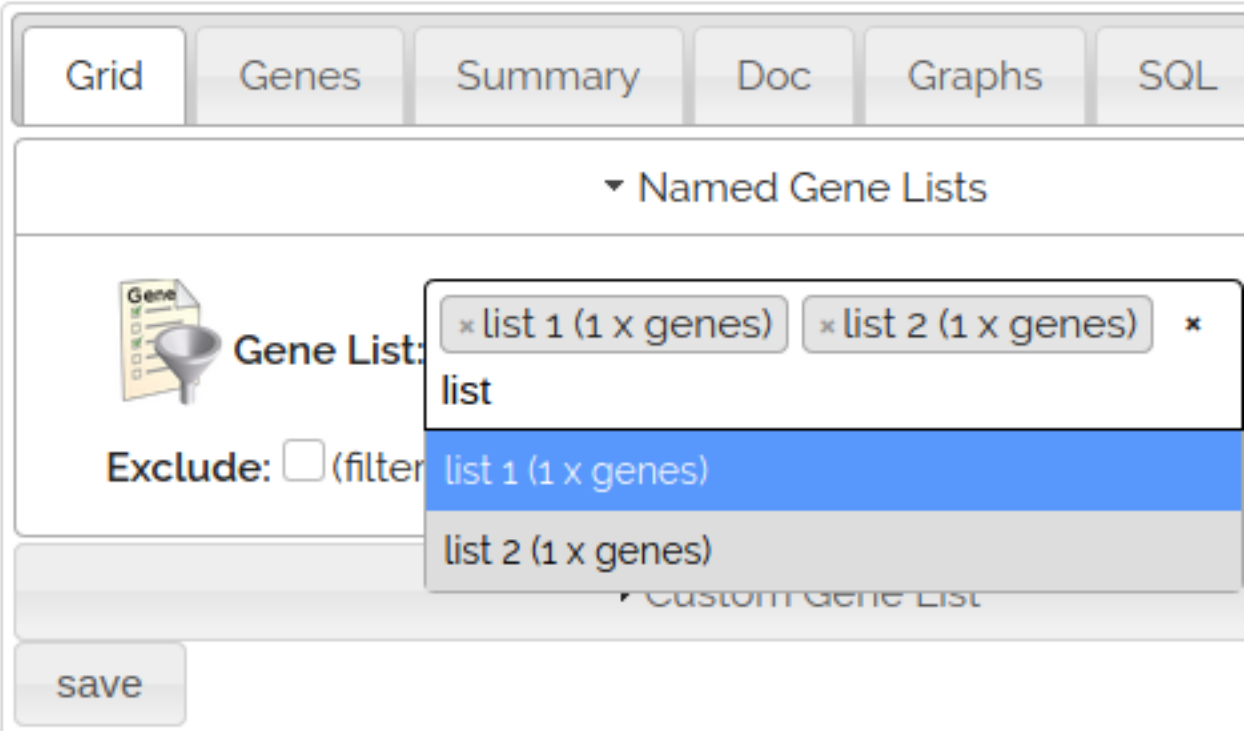
4.2.4 Filter

Construct your own filter based on column values. “All” means all lines must be met (AND), “Any” means any can be met (OR)

Search is case insensitive (except “in”). Some columns contain NULL (no value) which will not match anything. You may want to use “is null” to include or “is not null” to exclude them.

4.2.5 Gene List

Filter to a list of gene symbols.



Used **Named Gene Lists** to select existing *Gene Lists*. You can select multiple lists at a time.

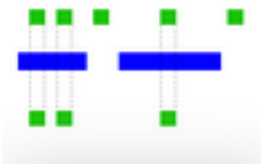
This node returns variants where ANY TRANSCRIPT matches the genes in the list, see *transcript choice*

Custom Gene List - Enter symbols directly, without having to create a gene list first.

PanelApp Panels - Displays a list of panels from Australia/England PanelApp which you can auto-complete to select.

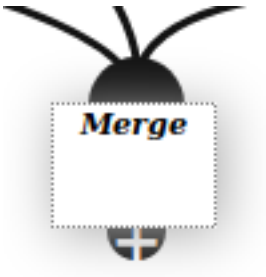
View the "Genes" tab to see which genes are being used by the filter.

4.2.6 Intervals Intersection



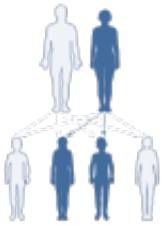
Filter based on intersection with genomic ranges (eg .bed files), a custom range (chrom: start-end) or a HGVS coordinate.

4.2.7 Merge



Merge variants from multiple sources

4.2.8 Mode of Inheritance



Uses known gene/disease associations from the Gene Curation Coalition (GenCC)

Disease ontology terms must be in MONDO as that is what is used by GenCC

If a sample is provided, with the "strict zygosity" option, that sample's zygosity will also be taken into account. For instance if a gene/disease mode of inheritance is "Autosomal recessive" then only homozygous variants in that gene will be included.

4.2.9 Phenotype

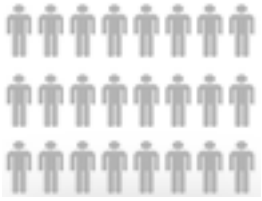
Filter to genes related ontology keywords (HPO, OMIM and MONDO). This is more lax than the Mode of Inheritance filter, as there are genes associated to a term but not definitively classified as disease causing.

You can autocomplete terms (multiple select) for exploratory analysis, however it is far better to actually *store the phenotypes against the patient*.

You can then select a patient to use those phenotypes (the patient must be assigned to a sample that is an ancestor to the pheno node)

View the "Genes" tab to see which genes are being used by the filter.

4.2.10 Population



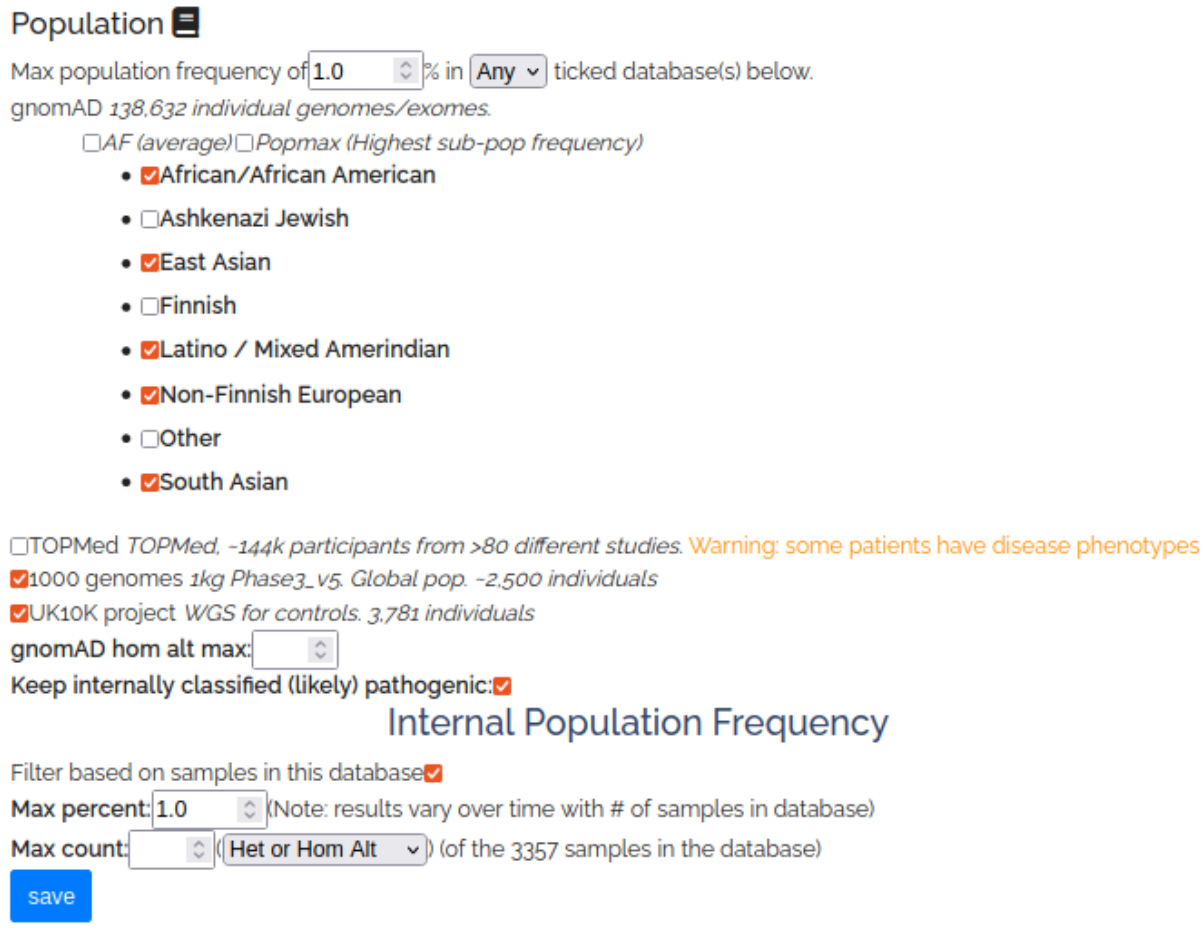
Most genetic diseases are rare (eg 1 in 10,000 people) so we know the disease-causing variant must also be rare. So when searching for disease causing variants, one of the first things to do is filter out variants that are common in the population.

This node filters variants by population frequency in public databases (gnomAD/TopMed/1KG/UK10K) or *internal frequency in this database*.

PopMax is the frequency of the highest sub-population (Note: gnomAD2 includes bottlenecked populations such as Finnish/Ashkenazi, while gnomADv3 excludes them)

Click "Pick individual gnomAD populations" to expand the selection to sub-populations (ancestry groups such as Europeans or East Asians).

You can also restrict to a max count (gnomAD hom alt or internal zygosity counts) which is useful to restrict to very rare variants (eg denovo)



Internal database frequency thresholds are critically dependent on what samples are in your database, most clinical

databases will be highly enriched for disease samples. If you have entered *patient phenotypes* you can see counts of disease terms on the patient page.

4.2.11 Tags

Filter variants to those that have been tagged. You can select multiple tags in the auto-complete. If no tag is selected, it filters to any tag.

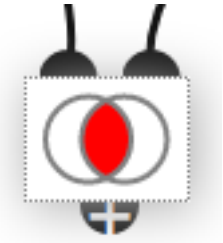
Tags From determines whether to filter to variants tagged just in this analysis ("Tags" column), or anywhere (both "Tags" and "Tags in other analyses" columns)

Parent Input - When set, the node has a top connector and filters the parent node's variants. If not set, the node is a source node and retrieves all tags.

The **Exclude** option removes variants with tags - this is most often used for filtering out artefacts (All tags).

The tags icon on the toolbar allows you to quickly see all tags in this analysis, without having to make and configure a tag node.

4.2.12 Venn



A filter based on set intersections between 2 parent nodes

4.2.13 Zygosity

Filter to an ancestor sample's zygosity. Multiple hit filters to variants where a minimum of 2 are present per gene.

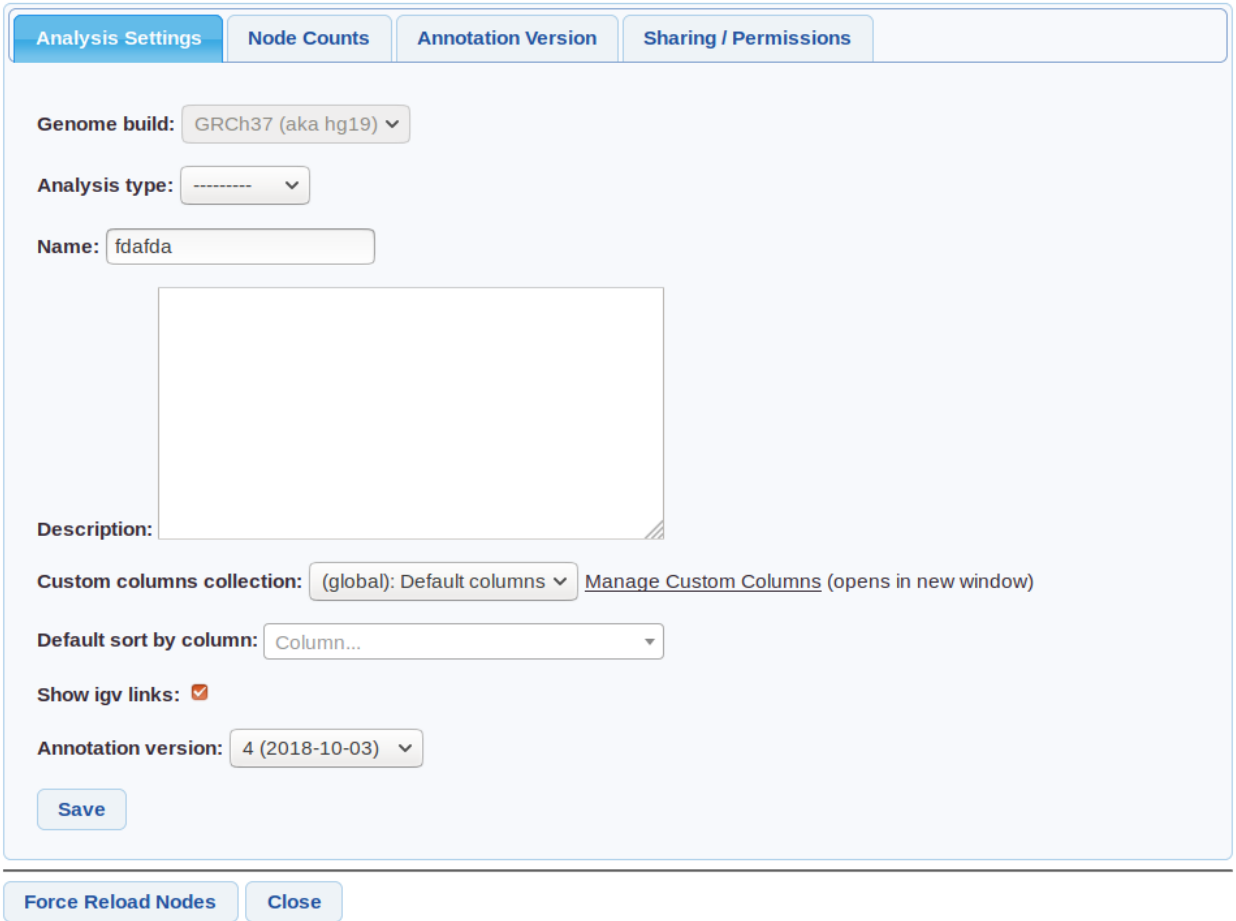
CHAPTER

FIVE

ANALYSIS - ADVANCED

5.1 Analysis settings

In an analysis click the Settings icon to open the analysis settings page.



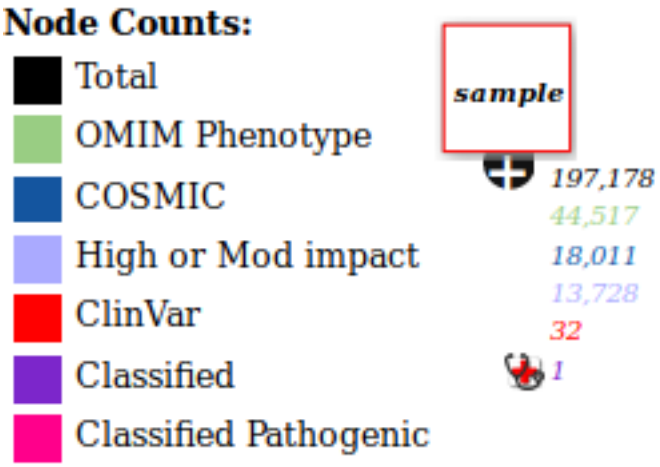
Analysis settings screenshot

- **Genome build** - Cannot be changed. Only data (eg VCF samples) from this build can be used in the analysis.
- **Analysis type** - One of (Singleton/Cohort/Trio/Pedigree) set at creation if using an auto-analysis.
- **Custom columns** - Columns to use - from *customise columns*. Default set in *user settings*
- **Default sort by column** - Can be used for example to make the grid always sort by gene.

- **Annotation Version** - The *Annotation Version* used.

5.2 Node Counts

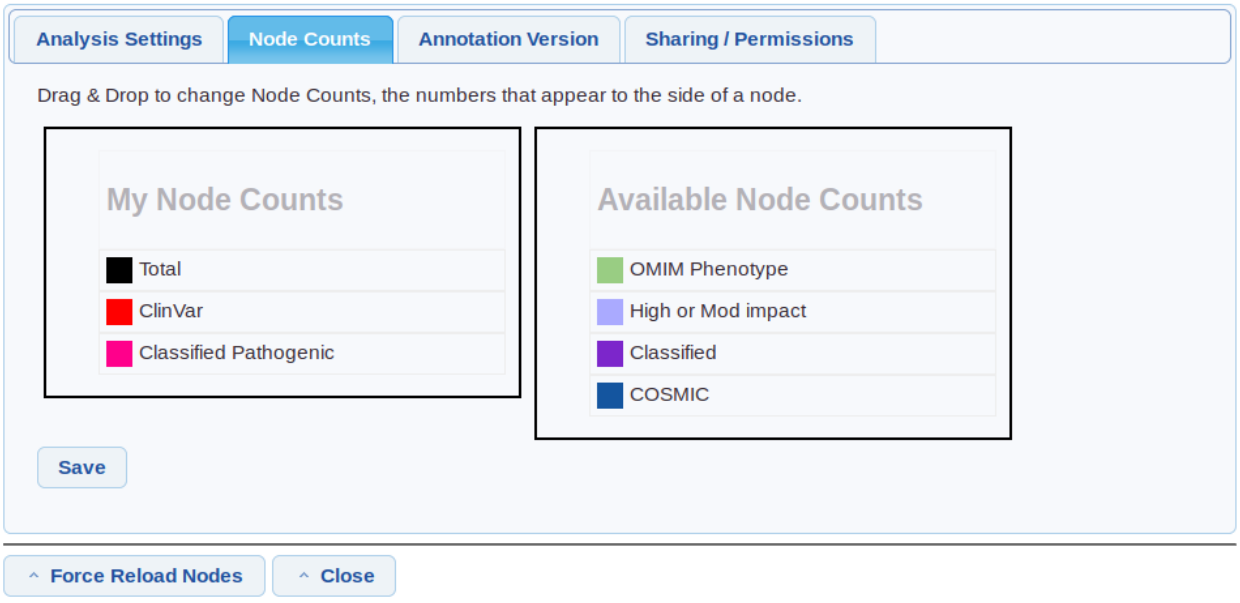
The numbers below a node are counts of variants that meet a certain criteria. The colours correspond to names in bottom left hand legend, eg in the image below, there are 32 ClinVar (Likely) Pathogenic variants in that node.



Node with counts

Click on a count to load the variants in the node that meet that criteria, eg clicking on the red 32 would just load the ClinVar variants.

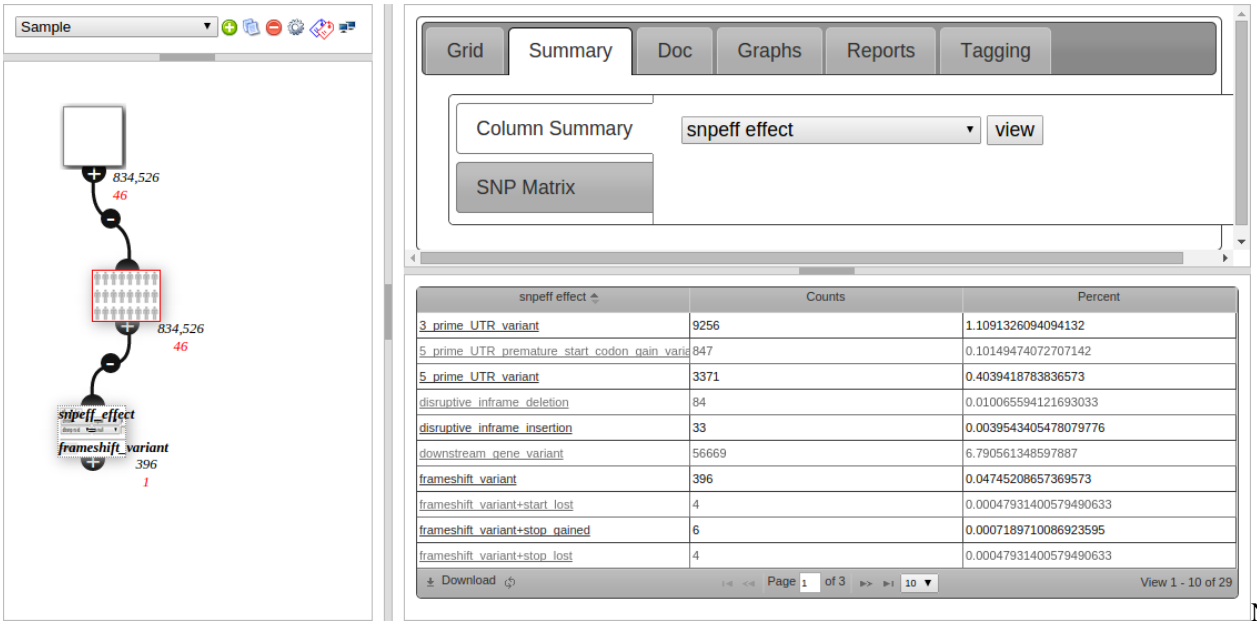
To edit which node counts are shown, open analysis settings, then select the "node counts" tab.



Settings/Node counts

Drag and drop the node counts to show/hide them and change the order.

5.3 Column Summary



Node Summary

The second tab (Summary) is used to view what values are in a column. Qualitative data is counted and shown in a grid, such as snpEFF Effect in the screenshot below:

Clicking on the link in the 1st column creates a child node filtering to that value. This is useful for getting an overview then drilling down into your data.

The screenshot shows 396 entries under “frameshift variant”, and the filter node created underneath the current (red bordered) node, which is configured to filter to snpeff_effect = frameshift variant, and also has 396 variants after filtering.

Quantative data (numbers, such as for the af_1kg column (1000 Genomes Alt Frequency)) is shown as a box-plot.

CHAPTER

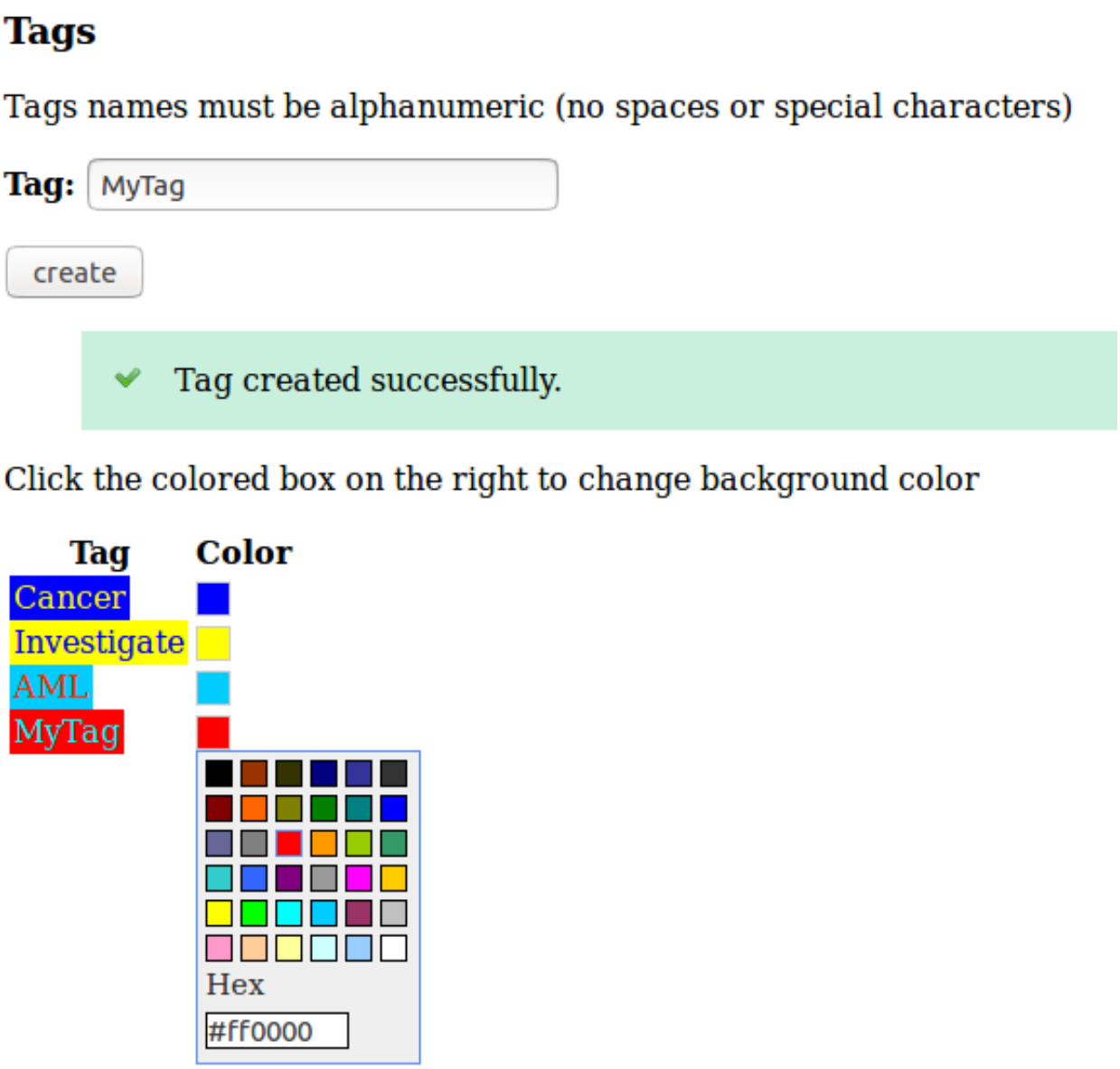
SIX

VARIANT TAGGING

A tag is a label (such as “Cancer” or “Investigate”) which you can use to label and track variants in an analysis.

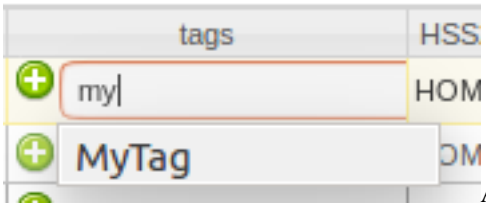
6.1 Create tags

Menu: **[settings] -> [tags]**



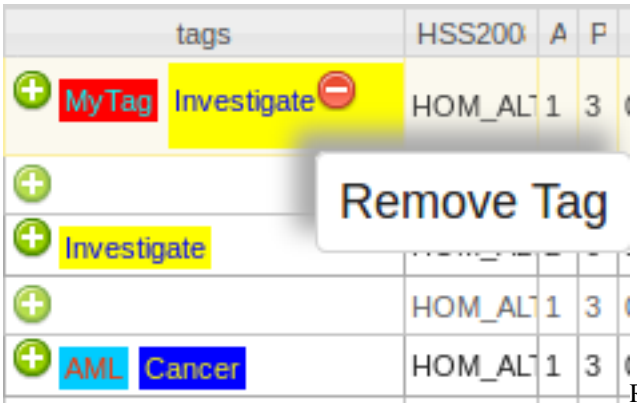
6.2 Tagging variants

In an analysis, click the Add icon in the "tags" column then auto-complete your tag.



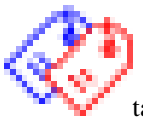
Adding a tag

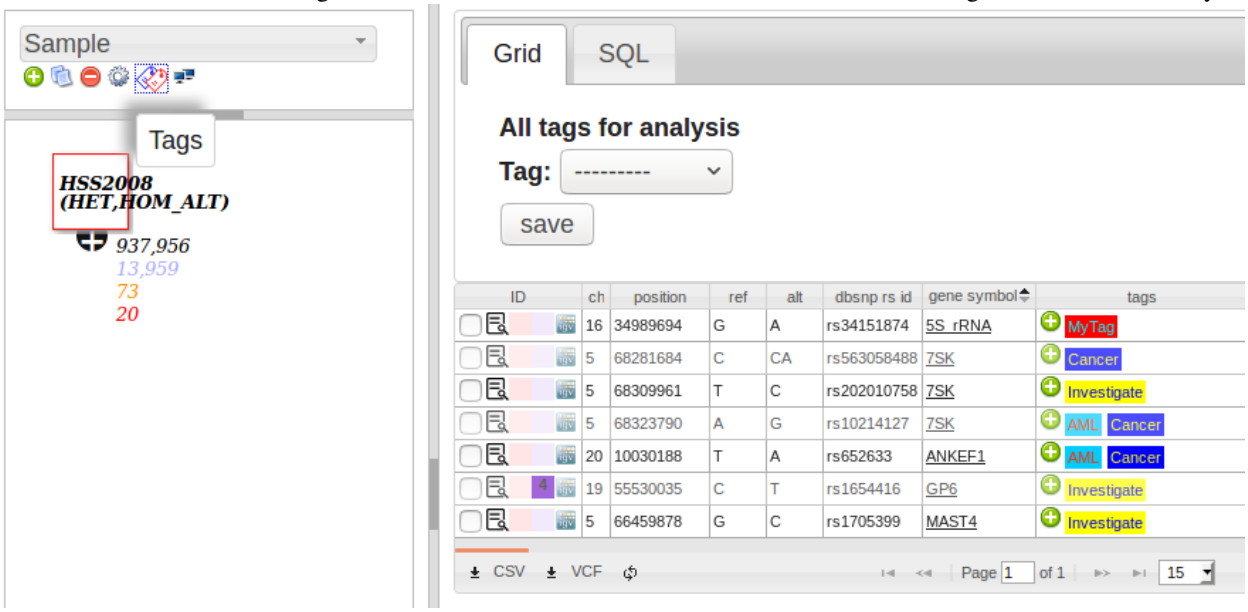
To remove a tag - clicking on the tag. The tag will grow in size, and a delete symbol will appear. Click it to remove the variant tag.



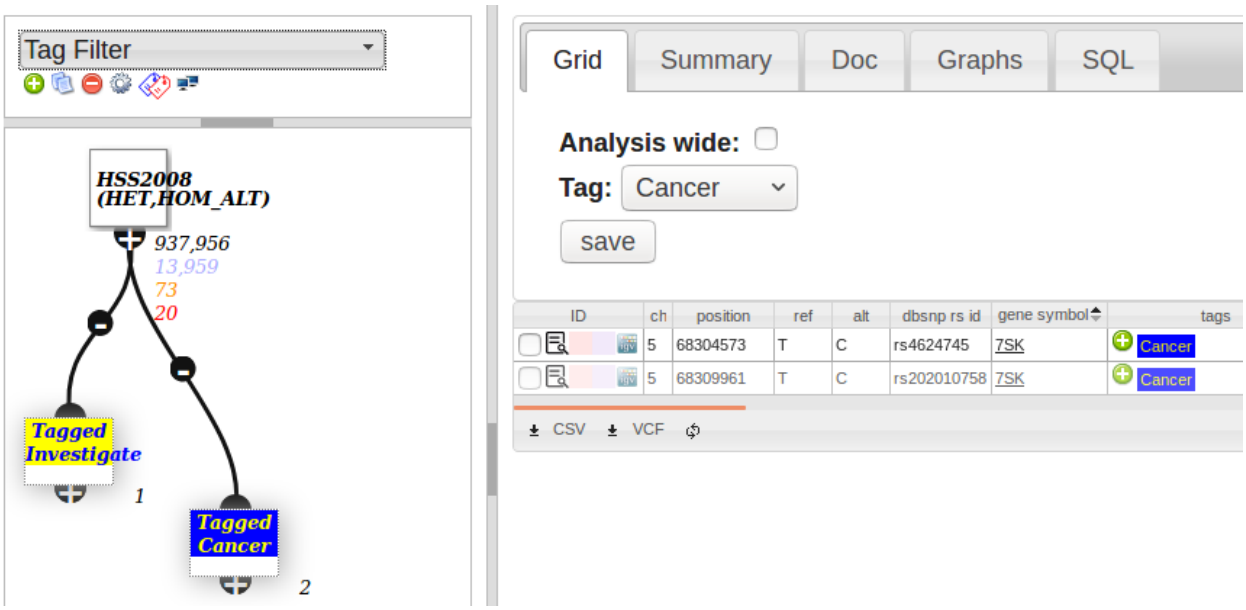
Removing a tag

6.3 Using tags

Click the  tag icon on the toolbar to view all Tags in an analysis



To filter to specific tags - add a tag node, and use it like any other node to filter variants to just those that have been tagged.



You can view all tagged variants on a page, via menu: **[analysis] -> [Tagged Variants]]**

CHAPTER

SEVEN

ANALYSIS CLASSIFICATION

Recommended workflow to create a *classification* from a variant in an analysis:

1. *Tag* the variant with the "RequiresClassification" tag.

ID	ch	position	ref	alt	dbsnp rs id	gene symbol	tags	HSS2008	AD	AF	DP	GQ	PL	HSS2009	AD	AF	D
2	12	49433599	T	G	rs147706410	KMT2D		HET	31	47.6	null	null	0	HET	24	44.4	n
2	12	49428694	T	C	rs146044282	KMT2D		HET	56	43.7	null	null	0	HET	56	49.1	n
	6	10410466	T	G	rs776792762	TFAP2A	RequiresClassification	HET	3	33.3	null	null	0	.	2	12.5	n

CSV VCF Page 1 of 1 15

1. Click the  tags button, then then "Classification" tab.
2. Select the sample, then click the **[classify]** button.

CHAPTER

EIGHT

ANALYSIS - TEMPLATES

Menu: **[Analysis]** -> **[Templates]**

8.1 Overview

The fastest and easiest way to run an analysis is to apply a pre-defined analysis template to your sample, trio or cohort. This allows you to quickly run the same analysis over different data without needing to build or edit analysis settings.

VariantGrid comes with a number of pre-configured analyses templates - all of which can be modified by the database admin as required. In addition, users can build their own templates using the template wizard. A template is built in the node editor in the template wizard the same way as a normal analysis, however, there is an option to configure the sample, trio and/or cohort fields as 'analysis variables', allowing these fields to be set from new data each time the template is applied.

To see the analysis templates currently available in your installation go to **[Analysis]** -> **[Templates]**.

8.2 Running Analysis templates

You can create an analysis from an existing template using the 'Create from Template' button in the "Analysis" section at the bottom of the Sample, VCF, Cohort, Trio and Pedigree pages. Each analysis template has an expected input type (sample, trio or cohort). Only templates that match the data type on that page are shown. For example, trio templates will only be displayed on the trio page, but not on a sample page and vice versa.

When an analysis has been created from a template, a 'Template Run' tab will be displayed in the analysis settings window. This will record a list of the variables that were used to generate the analysis.

If the template includes downstream nodes that are dependent on sample or patient-related inputs these nodes will be updated accordingly. This will occur in the following circumstances:

- Zygosity node - the input sample will be used as the zygosity sample
- Gene list node using a sample gene list - the active gene list for that sample will be applied
- Phenotype node using the patient phenotype - the phenotype terms will be updated to match the phenotype of the patient linked to the sample (if available)

The analysis can be modified as usual.

An analysis template is created for a particular genome build, but will run without error on any build provided build-specific data is not required in the analysis. For example, a GRCh37 analysis template containing a Genomic Intersection node bed file will not run as the GRCh37 file can't be applied to GRCh38 data. Note that care should be taken to validate that any filter settings used are applicable to the non-native build.

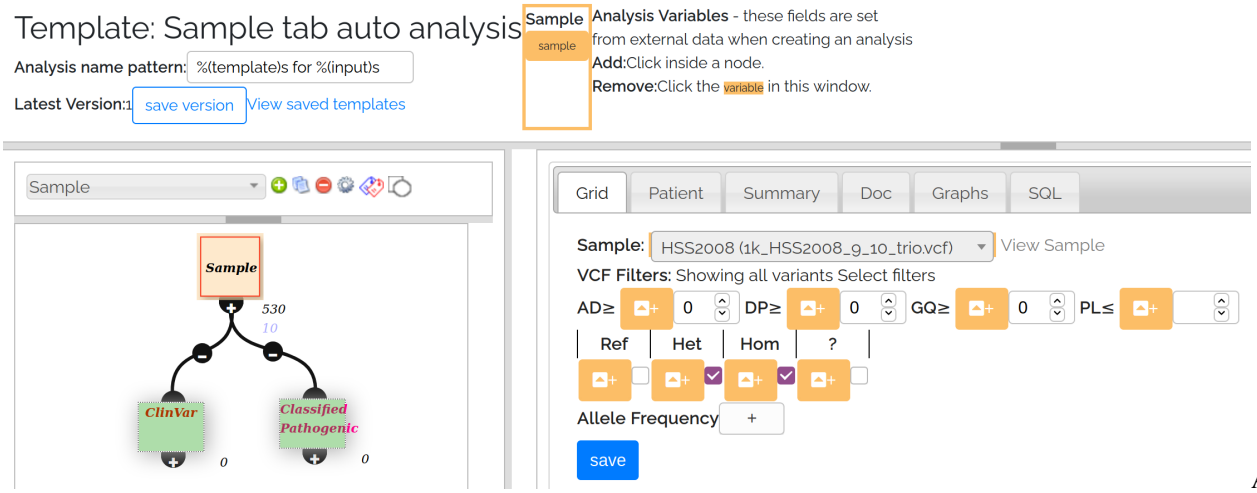
8.3 Creating Analysis Templates

There are two methods available to create analysis templates:

- Create a new (blank) template from the Analysis templates page **[Analysis]** -> **[Templates]** In this method you will need to manually set the source node sample field(s) as an analysis variable(s).
- Copy an existing analysis. **[Analysis]** -> **[Analysis settings]** (cog icon) -> **[Create Template]** tab -> **[Create Template from this analysis]** button. This method will automatically set the source node as an analysis variable.

To save a template, click the **[Save version]** button on the top bar. Templates must be saved before they can be used.

The screenshot below shows an analysis template in the analysis wizard window. Nodes colored orange contain analysis variables, which also appear in the top bar. Green nodes are 'output nodes' representing the filtered variants of interest.



Analysis Template screenshot

Setting an analysis variable: Open a node, then in the node editor click the orange button next to a field to make it an analysis variable. This will make the widget unselectable, and add the field to the top bar. In the example above the sample field has been set as an analysis variable. Currently only the sample fields on the sample, trio and cohort nodes can be set as analysis variables.

To remove an analysis variable, click on the field in the top bar.

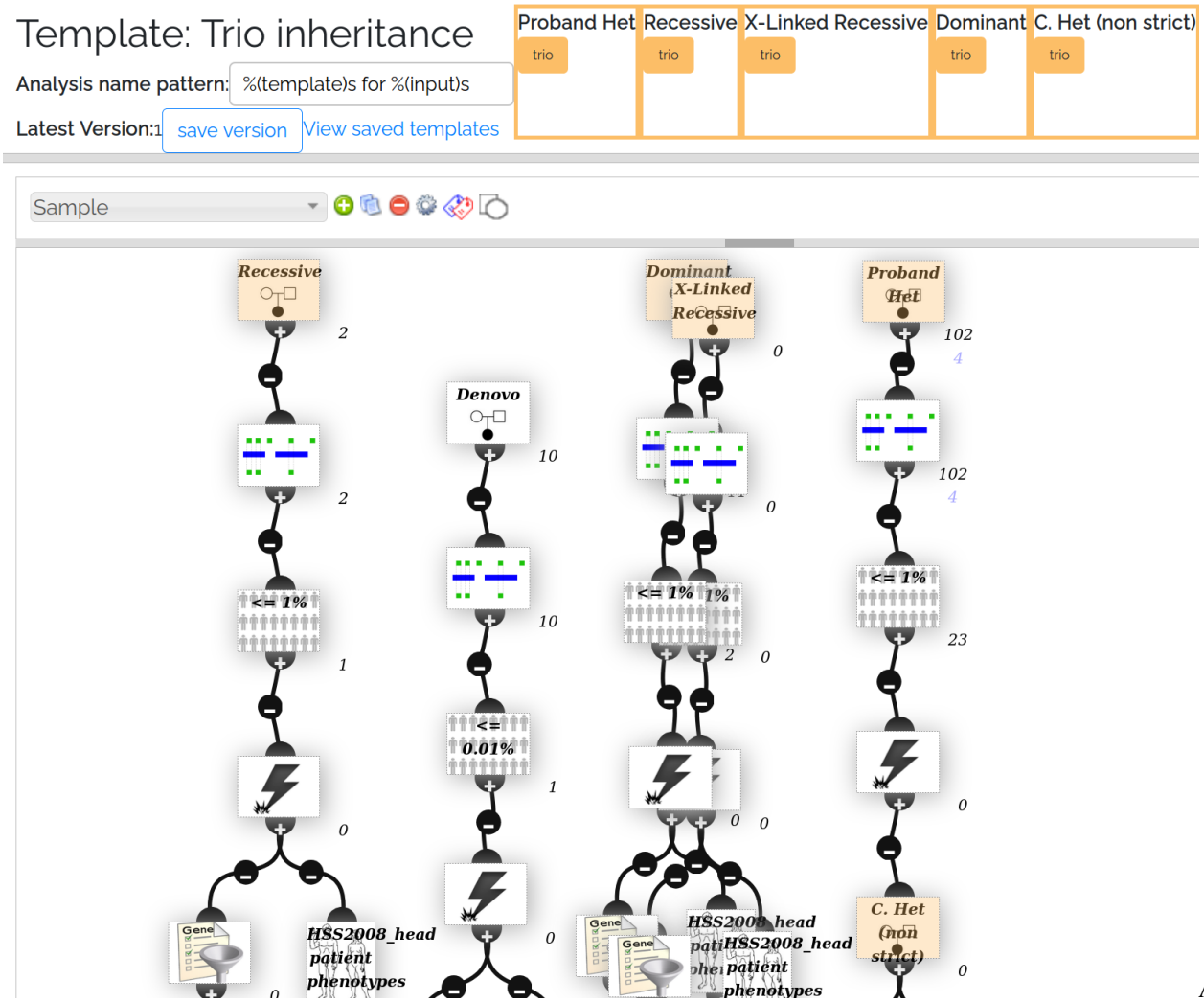
Setting an output node: To define an output node, click on the node and select the **[doc]** tab. Make sure the node has a good, unique name then select the **[output node]** check box and save. This will turn the node green indicating it is set as an output node.

8.3.1 Handling configuration failure

Sometimes parts of an analysis may not make sense depending on the input data. For instance in Trios, whether the parents are affected determines whether you want to use Dominant or X-linked inheritance model filters.

When an analysis is run, nodes run internal checks to make sure they are configured correctly, so for instance a TrioNode configured to "Dominant" on a trio with unaffected parents will be invalid (node and all descendants will error + flash red)

So to handle this, build all the filters in the template, then for nodes that you expect to sometimes error out due to configuration, go to the Node Editor **[Doc]** tab -> **[Hide node and descendants upon template configuration error]**



Analysis Template for Trio inheritance - in the template 2 filters overlap, but in the generated analysis only 1 will be shown

8.3.2 Tips and tricks

- In the Trio inheritance screenshot above, note the top right node is a TrioNode configured to "Proband HET". If you were building this analysis by hand, you might use a HET SampleNode, however this would then require you to have an anaylsis variable of type "Sample" (which we'd be unable to set via the Trio page)
- Node editors hide options based on data (eg GeneListNode will not allow you to select "sample gene list" if samples do not have one) so configure the template using data that is as similar as possible to what you intend to use.

8.3.3 Configuring where templates are shown

You can further configure how/where templates are shown (currently admin only)

- appears_in_autocomplete (default=True)
- appears_in_links (default=False)
- requires_sample_somatic (default=False)
- requires_sample_gene_list (default=False)

CHAPTER

NINE

KARYOMAPPING

9.1 Background

We handle the simpler case of a *Trio* with an affected child (ie proband/mother/father).

"In phase" implies that the allele from a parent is the same as that in the affected child

Variants are assigned to the following bins

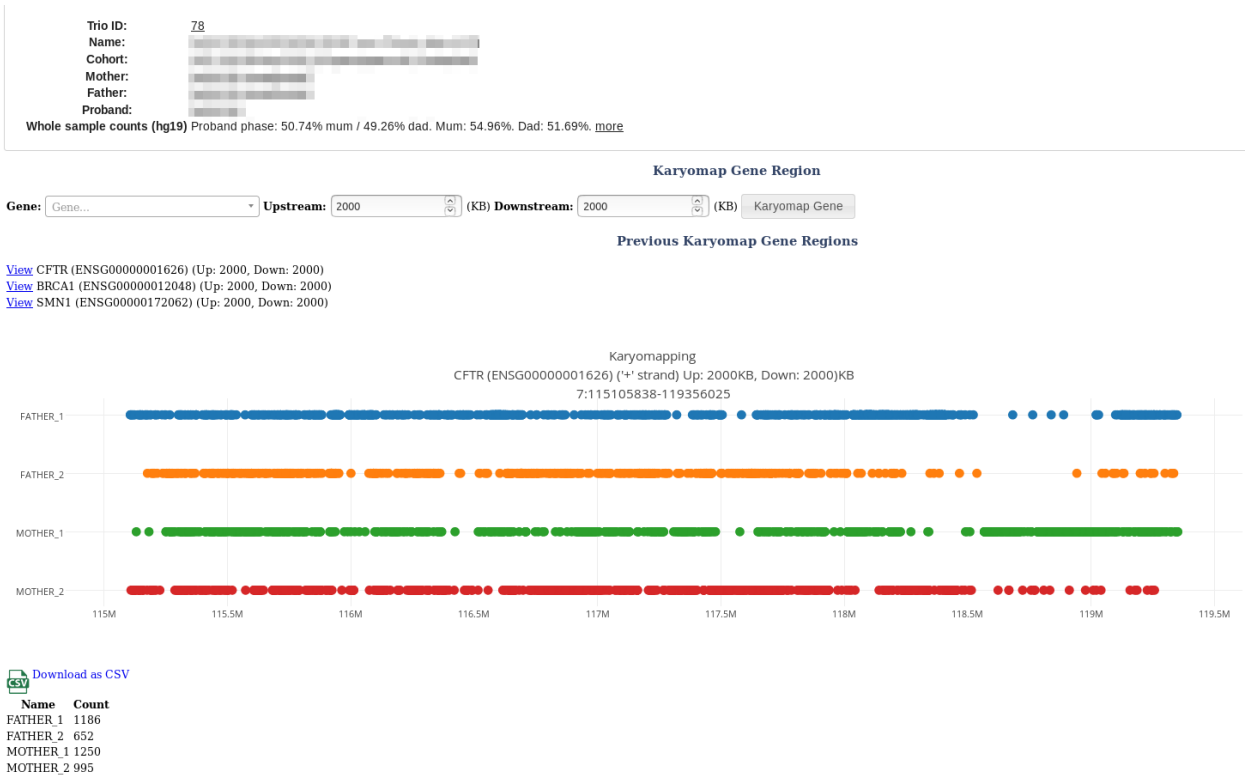
F1ALT: Paternally inherited, in phase with affected child, ALT variant. F1REF: Paternally inherited, in phase with affected child, REF variant. F2ALT: Paternally inherited, out of phase with affected child, ALT variant. F2REF: Paternally inherited, out of phase with affected child, REF variant.

And vice versa for the mother. The only variants that fall into each of these situations are:

9.2 Gene analysis

Menu: **[analysis]** -> **[karyomapping]**

Enter a gene name and click **[Karyomap Gene]** button.



9.3 Genome-wide analysis

A genome wide karyomap count is performed when you create a trio. This is useful for finding sample mixups.

This is summarised as *Proband phase: 50.74% mum / 49.26% dad. Mum: 54.96%. Dad: 51.69%.* and is visible on the gene analysis screenshot above and the *Trio* page.

Proband phase shows the child's marker percentage from each parent. Mum%/Dad% = Percent of parent markers that are in phase in proband.

Here are some examples for various Trios:

As a rough rule, you'd expect a minimum of 40% for an actual child.

CHAPTER

TEN

ANNOTATION DETAILS

Annotation refers to all of the information about a variant, it is made from different components, including:

Variant-level annotation: Information specific to a base change. Examples include computational predictions and effects, and existing database entries (such as population frequency for the variant)

Gene-level annotation: Information about the gene (from RefSeq/Ensembl + other sources), matched from the variant's assigned transcript_id.

ClinVar: Clinical variant classifications from ClinVar

To see a description of each field, use menu: **[annotation] -> [descriptions]**

Annotation is shown on the variant details page, and in an *analysis*, where it is used in filters and shown on the grid (see *customise columns*)

10.1 Variant Level Annotation

The first time we see a variant, it is annotated by Ensembl Variant Effect Predictor (VEP) and then cached in the database.

VEP calculates the effects for each transcript overlapping a variant, then picks a representative transcript - this is what is used for filtering in an *analysis* and shown in the grid.

10.2 Annotation Versions

Each annotation component above is versioned and can be upgraded separately by the site administrator. To see the versions via menu: **[annotation] -> [versions]**

VariantGrid can store multiple annotation versions, which allows us to load historical analyses which return the same results as when they were first analysed, as well as updating from new sources regularly.

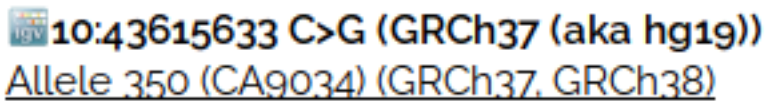
CHAPTER

ELEVEN

VARIANT DETAILS

This page shows the annotation and other information about a variant.

The top of the page has an IGV link, and a link to the allele for this variant:



An allele is genome build independent - ie hg19 and hg38 variants for same change point to same allele. The ID (CA9034) is from the ClinGen Allele Registry

11.1 Classifications

ID	HGVS	Clinical Significance	Condition	Curated Date	Flags
My lab / vc0042	NM_000130.4(F5):c.1601G>A	Benign (1)		2019-08-06	

Variant Details - Classification section

This shows internal *classifications* for an allele (may have been classified against a different genome build)

The far right column contains *Classification Flags*

11.2 Transcripts

Variant annotation is calculated for each transcripts overlapping a variant. You can select each of the different transcripts to change which is being displayed. A transcript can be labelled as Representative (most damaging for variant shown on analysis grid) or canonical (transcript chosen for gene by RefSeq/Ensembl)

11.3 Samples

At the bottom of the page is a grid of samples that contain the variant (and the zygosity and read information). Only samples you have permissions to view are shown, but a warning will be shown informing you that samples you don't have permission to see exist.

CHAPTER

TWELVE

TRANSCRIPT CHOICE

Variants are annotated with multiple transcripts, which can give different results.

Shown below is a variant that overlaps with two different genes (ANKEF1 and SNAP25-AS1) with many transcripts:

	Gene	RefSeq	Ensembl	HGVS c.	HGVS p.	Molecular Consequence	Impact	Properties
○	ANKEF1		ENST00000378392.1	ENST00000378392.1:c.971T>A	ENSP00000367644.1:p.Leu324Gln	missense variant	MODERATE	
○	SNAP25-AS1		ENST00000421143.2	ENST00000421143.2:n.235-22892A>T	-	intron variant & non coding transcript variant	MODIFIER	
○	ANKEF1		ENST00000488991.1	ENST00000488991.1:n.1278T>A	-	non coding transcript exon variant	MODIFIER	
○	SNAP25-AS1		ENST00000603542.1	ENST00000603542.1:n.748-22892A>T	-	intron variant & non coding transcript variant	MODIFIER	can.
○	ANKEF1	NM_001303472.1		NM_001303472.1(ANKEF1):c.404T>A	-	?		
○	ANKEF1	NM_022096.4		NM_022096.4(ANKEF1):c.971T>A	-	?		
○	ANKEF1	NM_198798.1		NM_198798.1(ANKEF1):c.971T>A	-	?		
○	SNAP25-AS1	NR_040710.1		NR_040710.1(SNAP25-AS1):c.500-22892A>T	-	?		
●	ANKEF1		ENST00000378380.3	ENST00000378380.3:c.971T>A	ENSP00000367631.3:p.Leu324Gln	missense variant	MODERATE	rep. can.

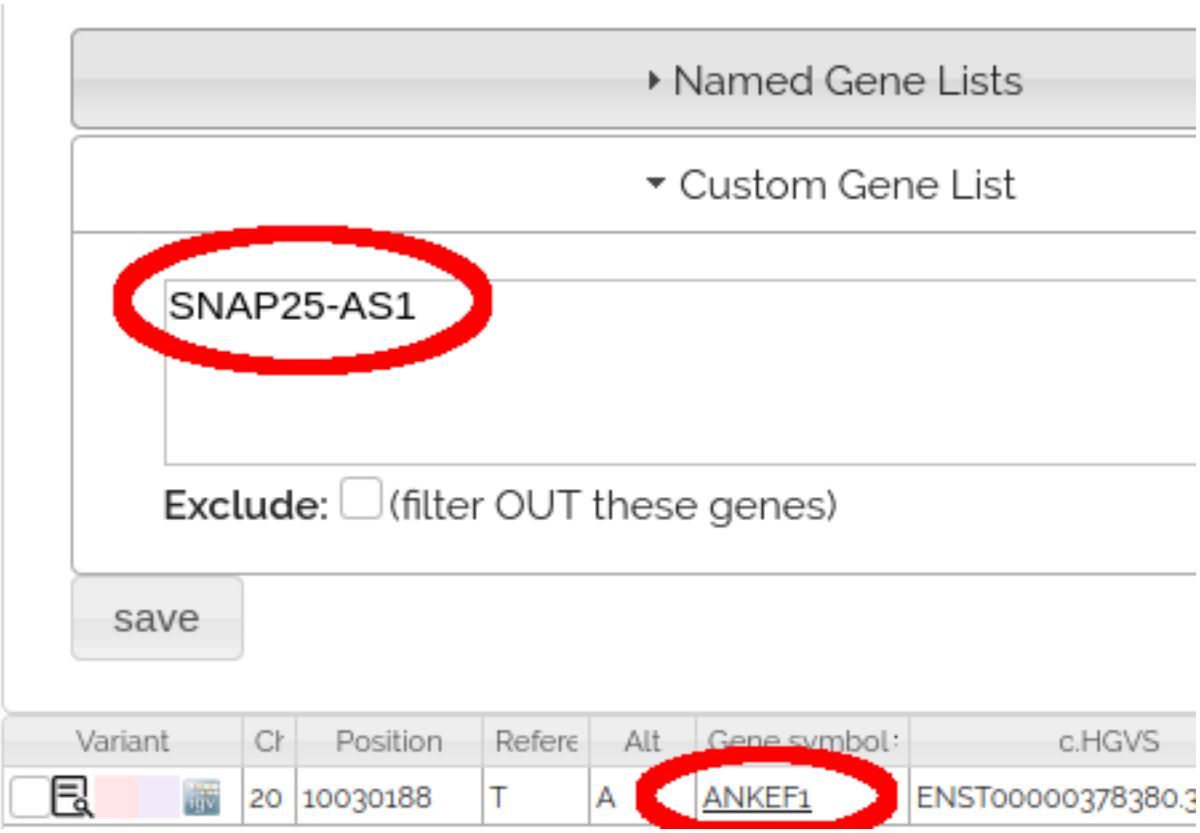
12.1 Analysis transcripts

We only want 1 row per variant in an *analysis grid*, so a single **representative transcript** is chosen to be displayed and filtered (see below)

You can see annotation for all of the transcripts by clicking on the 1st column in the grid to open variant details

Most *analysis nodes* filter on fields from the representative transcript shown on the grid, so representative transcript choice can affect analysis results.

The GeneList Node returns variants where ANY TRANSCRIPT matches genes in the gene list, not just the representative transcript. For example, the variant at the top of this page has ANKEF1 as the representative transcript, but is returned when searching for SNAP25-AS1:



This ensures no variants are lost in gene list filters due to transcript choice, but leads to the unexpected behavior that variants may have gene names not in the gene list.

12.2 Representative Transcript

Chosen via VEP pick algorithm:

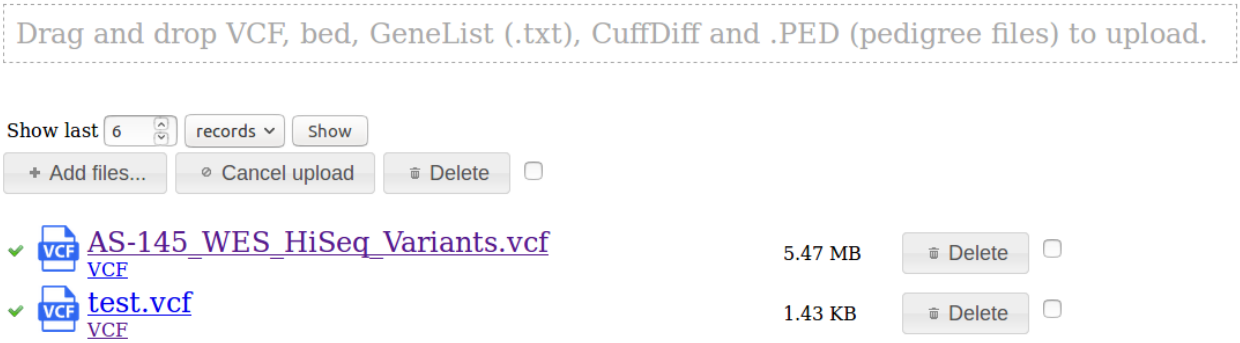
1. Canonical status of transcript
2. APPRIS isoform annotation
3. Transcript support level
4. Biotype of transcript ("protein_coding" preferred)
5. CCDS status of transcript
6. consequence rank according to this table
7. Translated, transcript or feature length (longer preferred)
8. MANE transcript status

CHAPTER

THIRTEEN

UPLOADING DATA

Menu: **[data] -> [upload]**



You can either drag & drop files onto the page, or by selecting the **[Add Files]** button.

After the file has been transferred to the server, a spinning icon () will appear as the file is processed. The large link (eg "AS-145_WES_HiSeq_Variants.vcf") takes you to the import processing page if you'd like to monitor the progress.

Once it has been successfully imported, a link will appear beneath the file (eg the "VCF" links above) allowing you to jump to the data page for this file.

CHAPTER

FOURTEEN

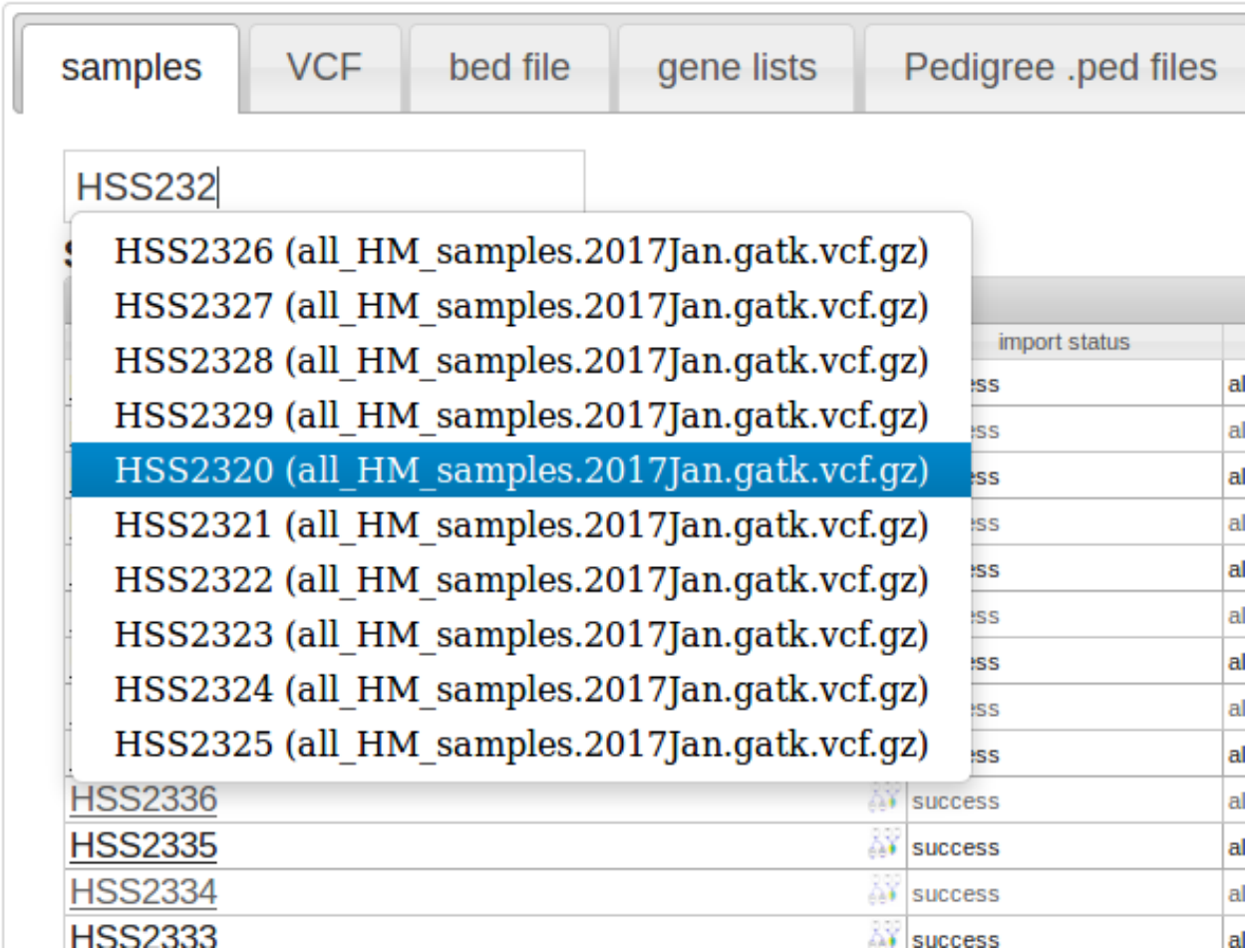
MANAGING DATA

Menu: **[data]**

The data page displays all of your uploaded data such as (VCFs, Bed files, Pedigree Files etc)

Data is displayed in grids, with each data type in a separate tab.

You can enter parts of the name into an autocomplete search box to quickly find your files:



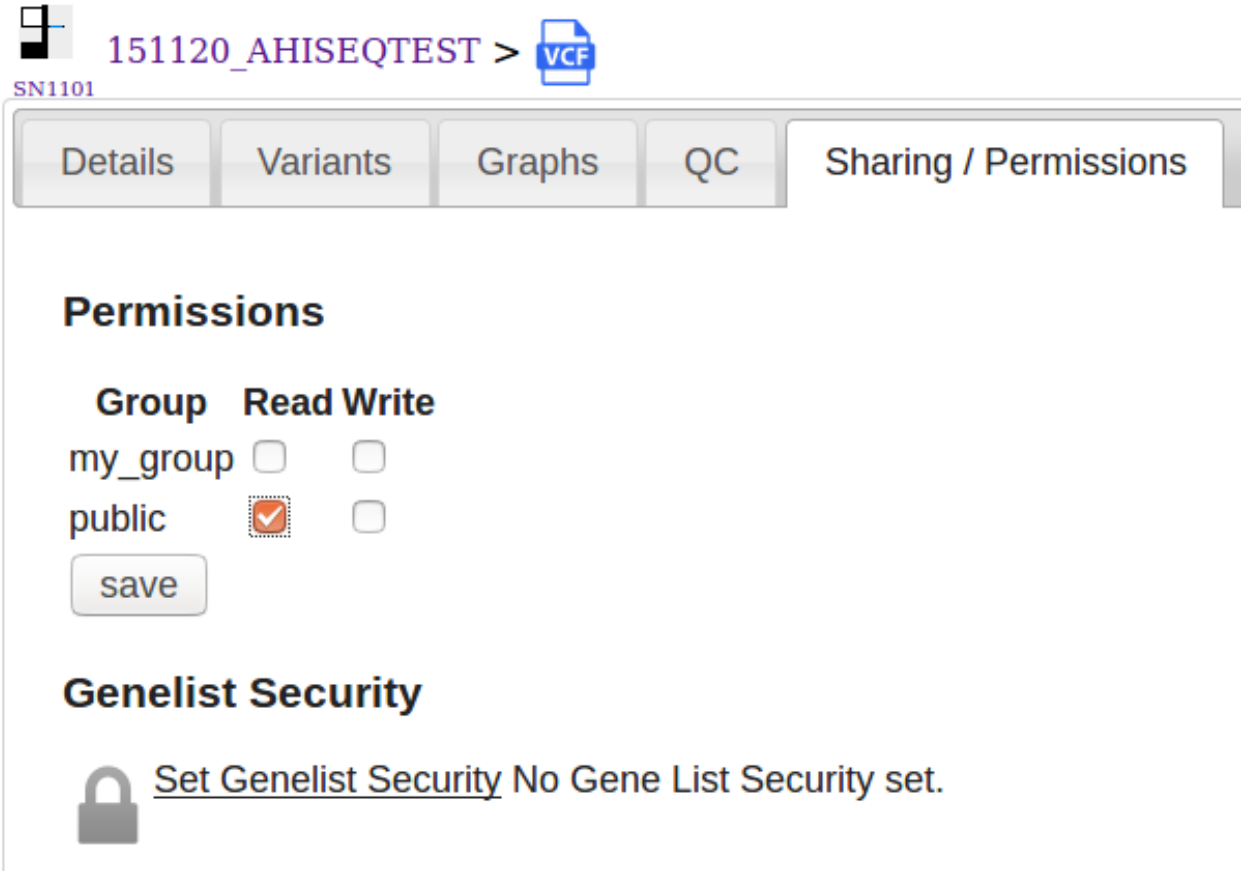
Click the link on the grid to view the file details page.

14.1 Sharing data

Users belong to groups (see *user settings*) that can share data. Ticking the **Show Group Data** checkbox will show this on a grid.

By default, you automatically share data (read-only) with your group.

To change data permissions, click the **[Data/Sharing]** tab:



logged_in_users is a special group - and means everyone who has a VariantGrid account.

14.1.1 HGVS

We use PyHGVS library for parsing HGVS names, which supports 'c.', 'n.' and '.p'.

CHAPTER

FIFTEEN

SOMATIC DATA

Somatic VCFs detected as somatic only (tumor minus normal) are analysed for mutational signatures

15.1 Allele Frequency

If the VCF contains an "AF" value, we will use that. Otherwise we will

We do not import the AF value from the VCF, but instead [normalize](../vcf_samples.md#VCF Normalization) the data then recalculate AF to be `AD / sum(AD for all variants at locus)`

In an analysis, nodes that represent one or more VCF samples (Sample, Cohort and Trio nodes) can filter by allele frequency.

Click the "+" button to add more sliders for AF ranges (between 0 and 100%) you will allow through (AF in any of the slider ranges will be allowed through)

For nodes with multiple sample (Cohort and Trio nodes):

all: all samples must have AF within the range sliders

any: at least one sample has AF within the range sliders

CHAPTER

SIXTEEN

MUTATIONAL SIGNATURES

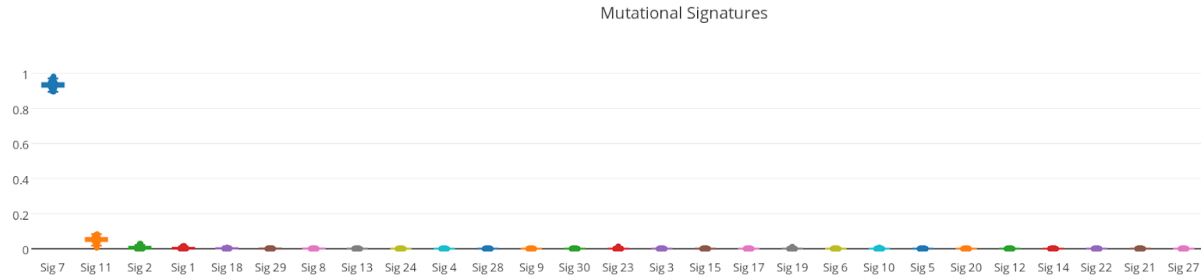
The type of somatic mutations present in a cancer sample can provide insights into the underlying molecular mechanisms driving oncogenesis. For example, cancer caused by tobacco exposure will result in an increased number of C>A transversions compared to cancers unrelated to tobacco. The advent of large cancer datasets has identified at least 21 conserved mutation signatures indicative of exposure or defective DNA damage repair mechanisms. For further details see Signatures of mutational processes in human cancer, Alexandrov et al 2013

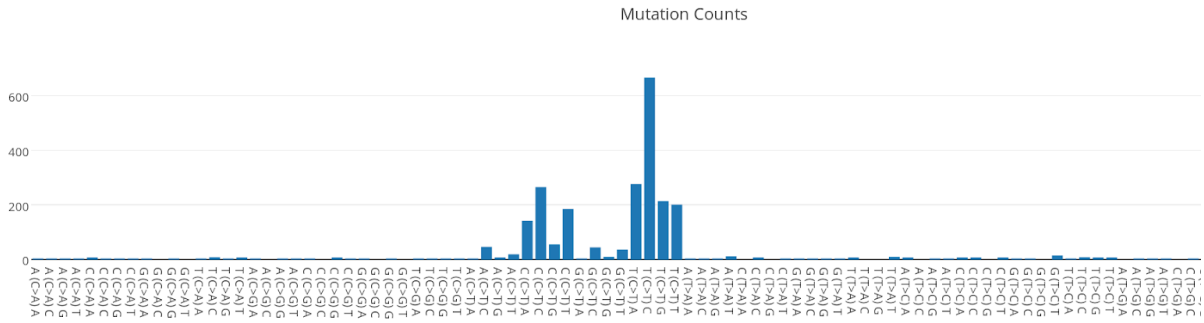
VariantGrid will automatically run mutation signature detection at vcf upload if the vcf is detected as a somatic only (germline subtracted) sample. VariantGrid recognises a sample as 'somatic only' based on information provided in the vcf header. Your VariantGrid administrator will need to setup the VCFSource object config to enable this functionality. It is not possible to manually run mutation signature analysis in VariantGrid once the vcf has been uploaded.

To view a mutation signature report go to:Menu: **[data]** -> Sort samples grid by "Mutational Signature" column -> Click on entry.Or click on the link in the "Mutational Signatures" at the bottom of the sample page.

In the example report below, the top graph indicates the percent composition of each mutation signature assessed. The bottom graph illustrates the frequency of each mutation type. In this example, the predominant mutation signature is found to be associated with UV damage.

Sample	161130HamishScott_somvar_combined.snvs.HC.vcf (161130HamishScott_somvar_combined.snvs.HC.vcf)
Summary	Sig 7: 93.51%, Sig 11: 5.23%
# snps	2375
Iterations	100
Sampling Fraction	0.8
Minimisation Strategy	Least Squares





Signature	Mean	Std	Cancer Types	Proposed aetiology
Sig 7	0.935142778421159	0.01660961417180388	Found predominantly in skin cancers and in cancers of the lip categorized as head and neck or oral squamous cancers.	Based on its prevalence in ultraviolet exposed areas and the similarity of the mutational pattern to that observed in experimental systems exposed to ultraviolet light Signature 7 is likely due to ultraviolet light exposure.

Thanks to Paul Wang from the ACRF Cancer Genomics Facility for the code.

CHAPTER

SEVENTEEN

VCF / SAMPLES

17.1 VCF import

Variants are normalized (see below) upon import. We only import variants, filters and genotypes (we don't use INFO as we do our own annotations)

The VCF format can vary a lot, we have tested VCFs from the following variant callers:

- GATK
- FreeBayes

Each sample is assigned a "variants type" of *Unknown*, *Germline*, *Mixed* (single sample) or *Somatic only* (tumor minus normal).

This is determined by looking at the "source" entry in the VCF header, and matching it to an entry in **VCFSource** object (setup by your administrator)

Samples with variants type of_somatic only_ are checked for mutational signatures

17.2 Multi-sample VCFs

Multi-sample VCF files combined using bam files record the genotype for all samples at each variant position.

This allows you to differentiate between reference calls and no coverage - and is extremely important for Trios so that you can make correct calls about inheritance and denovo variants

You must use bam files, to re-call the genotypes for each position.

Consider 3 VCF files:

There's no way to tell if a variant not being present in a single sample VCF is due to having the reference allele or no coverage.

Merging just the VCFs (without supplying the bams) will give the genotypes of:

If you merge them using GATK/Picard using bam files - the caller will re-examine the reads over the locus, and make the genotype call.

Thus, if both parents had reference bases, the calls would be:

And you can be confident that it is a denovo variant, rather than just lacking coverage in one of the parent samples.

17.3 VCF Normalization

We Decompose and Normalise variants using VT during import, so variants from different VCF files have a consistent representation.

If any variants were altered during an import, a warning appears on the VCF and Sample pages, allowing you to examine the changes.

You can search on an unnormalized variant, and it will take you to the normalized *variant's details page*. This page lists all VCF records normalized to that variant coordinate.

CHAPTER

EIGHTEEN

RELATEDNESS

Relatedness and ancestry is calculated using Somalier by Brent Pedersen + team, for more information on details, please read the paper

When a VCF is first uploaded, Somalier records each sample's genotype calls at ~18k coding sequence sites (sites have independent genetic linkage and approx 0.5 allele frequency in gnomAD)

With every new VCF uploaded, Somalier compare all samples against each other to generate pair-wise relatedness cooefficient scores.

18.1 VCF relatedness

Somalier generates a relatedness report which can be viewed in the **relate** tab on the VCF page (example report)

This shows a graph of samples relationships in the VCF. Sample to sample relatedness (ISBO vs ISB2) is plotted on the left. Related samples will locate in the top left and unrelated samples in the bottom right of the graph. Hover over a datapoint to see details of the pair-wise comparison. The Sample Depth Metrics plot on the right is used to display QC results. 4 commonly used visualisations are provided as quicklinks at the bottom of the graphs.

18.2 Sample relatedness

Samples with a minimum of 1000 shared HET, 1000 shared HOM and relatedness cooefficient >= 0.2 are displayed as a list at the bottom of the sample page.

Column Descriptions:

Interpreting relatedness:

- 1.00 = Identical twins/Clones/Duplicate sample
- 0.50 = Parent/Child/Full Sibling/Parent's Identical twin/Identical twin's child
- 0.35 = 3/4 sibling (e.g. same mother, fathers are siblings)
- 0.25 = Grandparent/Grandchild/Half-sibling/Uncle/Aunt/Niece/Nephew/Double-1st Cousin
- <0.20 = Distant relative (cousins) / Not related

CHAPTER

NINETEEN

ANCESTRY

Somalier creates an ancestry report which can be viewed on the **ancestry** tab of the VCF page example report). This feature is currently "experimental"

Samples are displayed on a PCA plot with individuals from the 1000 genomes project, which have labelled ancestries.

Somalier makes an ancestry prediction by comparing a sample with clusters from data with labelled ancestries.

The reported ancestry on the samples grid is the primary one and does not include admixture. A full breakdown of scores for all population groups can be found on the **ancestry** tab on the view sample page.

19.1 Implementation details

The amount of sites used will depend on a Sample's capture regions and sequencing depth (default min of 7). At least 1000 informative sites are required for robust calculation of the relatedness coefficient.

You can view how many sites Somalier used from a sample by going to the **ancestry** tab on the view sample page (under "Extract")

If a large number of unexpected samples are displayed as related, confirm the sample data type is an accepted input and that the sample has passed QC.

Comparisons work across genome builds and tissue types and can be used to compare RNA-seq, WES, bisulfite and WGS data.

Accepted samples:

- Multi-sample vcf: Ideal input
- Single sample vcf: Missing variants are assumed homozygous for reference allele.
- Tumour-normal vcf: Not recommended as no common sites due to germline subtraction

CHAPTER

TWENTY

SEARCH

Enter text into the search box in the top right hand corner and press enter or click Go.

This searches on the default build in your *user settings*.

If there is only one result, it automatically jumps to that page, otherwise it displays the results.

Click on the **Go** button without entering anything in the search box to visit the search page, where you can select which genome builds to search on.

Example inputs:

For HGVS, if no transcript version is provided, the most recent is used.

CHAPTER

TWENTYONE

ZYGOSITY COUNTS

VCFs with samples contain genotype calls (UNKNOWN/REF/HET/HOM ALT)

We store zygosity counts from for each variant for the samples in a VCF. This is used by the CohortNode to filter by zygosity and display the "hom count" and "het count" columns.

21.1 Global Counts

These counts are also stored globally - ie zygosity counts from a VCF can be added when it is uploaded, and subtracted if it is deleted.

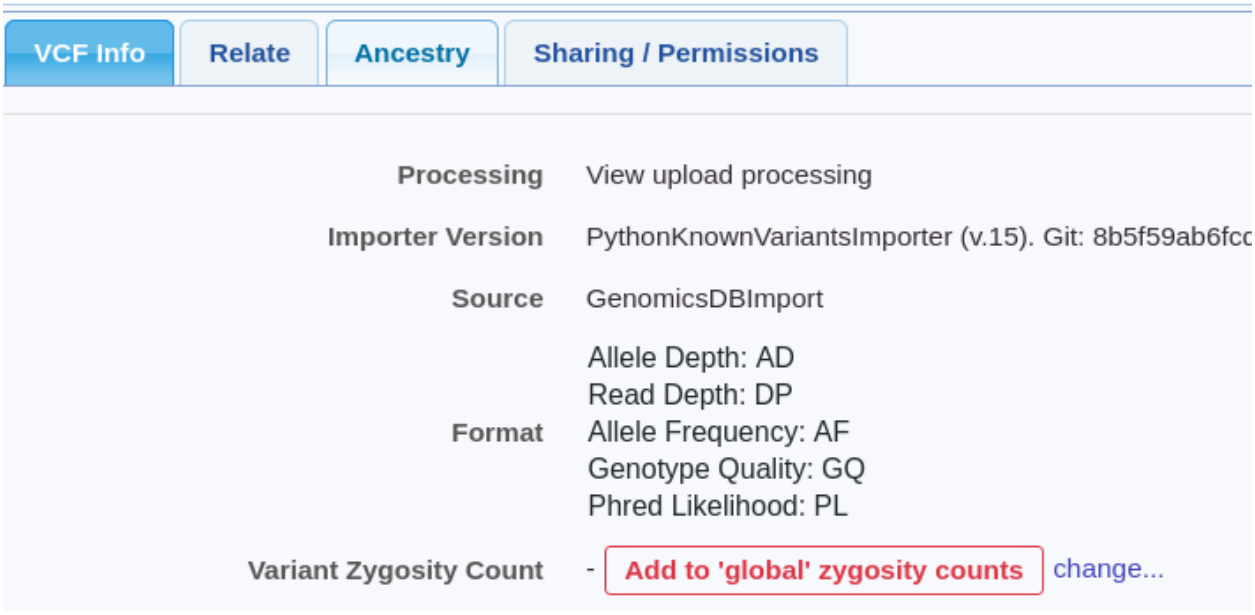
This is available on the grid as "database HOM count" and "database HET count" columns, and by the PopulationNode to "Filter based on samples in this database"

21.2 VCF configuration

An administrator can configure whether VCFs are added to the global count based on the VCF header or EnrichmentKit, for instance to ignore duplicate VCFs or only store germline samples.

You can see if a VCF is part of global zygosity counts by going to the VCF page, then the VCF Info tab, and the **Variant Zygosity Count** entry.

You can manually add/remove the VCF by clicking "change…" then hitting the button.


VCF Info
Relate
Ancestry
Sharing / Permissions
Processing
View upload processing
Importer Version
PythonKnownVariantsImporter (v.15). Git: 8b5f59ab6fcc
Source
GenomicsDBImport
Format
Allele Depth: AD
Read Depth: DP
Allele Frequency: AF
Genotype Quality: GQ
Phred Likelihood: PL
Variant Zygosity Count
- Add to 'global' zygosity counts
change...

CHAPTER

TWENTYTWO

GENES

22.1 Genes and symbols

It is worth separate the concepts of a Gene ID (eg ENSG00000179348) from a symbol (eg GATA2)

Ensembl Genes are versioned, eg the most recent version for GATA2 in GRCh37 is ENSG00000179348.7 and GRCh38 is ENSG00000179348.12

RefSeq genes are numbers without versions.

The symbol assigned to a gene can change over time (annoyingly, this is independent of the gene/transcript version). This is usually noticed across genome builds as the versions are often years apart.

22.2 RefSeq vs Ensembl

VariantGrid contains both Ensembl and RefSeq genes and transcripts, but a server can only be configured to run variant-level annotation (via VEP) for one.

You can classify against either, but on a server configured to use RefSeq, Ensembl transcripts will not have a molecular consequence or data for auto-population such as splicing calculations.

You can see what your system uses on the annotation page, by looking at "Gene Annotation Release"

22.3 Gene Annotation Release

A Gene Annotation Release is a snapshot of Gene IDs/versions and symbols - for instance "Ensembl v87" or "RefSeq v204"

This ensures our combination of symbols/gene+transcript versions match what is used by VEP, while allowing us to import new transcripts into the database (useful for resolving HGVS and interoperability between systems)

Each symbol in a gene list is mapped to a gene version in a Gene Annotation Release, so that filtering remains consistent over time, even if we later import new annotation which alters the symbol for a gene version.

You can see what gene versions and symbols are used by going to the genes page **[genes] -> [genes]**

22.4 Gene Annotation Grid

The data in the gene annotation grid can be explored using the OMIM quick filters that will filter to genes with corresponding OMIM data. Alternatively, use the search link to access the advanced filter.

Enter a gene symbol in the 'Jump to gene' dropdown or click on the gene symbol in the grid to navigate directly to a gene symbol page.

CHAPTER

TWENTYTHREE

GENE SYMBOL PAGE

The Gene Symbol shows annotation and internal data for a gene.

To see details of the genes IDs and transcripts associated with the gene symbol click on the RefSeq and Ensembl links at the top of the page.

23.1 Gene Annotation

Information on the page is combined from a wide-range of sources as follows:

- Aliases: A list of all gene symbol aliases. A warning is shown if the alias maps to multiple gene IDs.
- Summary: imported from RefSeq (if gene symbol linked to RefSeq gene)
- HGNC: Information derived from the HUGO Gene Nomenclature Committee based on the given gene symbol
- Uniprot: Information derived from the UniProt protein database
- Gene/Disease associations: ClinGen gene-disease assertions. Only available for ClinGen curated genes.
- gnomAD gene constraints: "The observed / expected (oe) number of loss-of-function variants in that gene. This a measure of how tolerant a gene is to loss-of-function. When a gene has a low oe value, it is under stronger selection for that class of variation than a gene with a higher value. For the interpretation of Mendelian diseases cases, we suggest using the upper bound of the oe CI < 0.35 as a threshold if needed. Ideally oe should be used as a continuous value rather than a cutoff and evaluating the oe 90% CI is a must." (extract from gnomAD)
- PanelApp: Gene panels from Geneomics England and Australia PanelApp websites. Note that PanelApp data is updated on a periodic basis. The date of last update is available in the annotations menu. Contact your VariantGrid administrator if a PanelApp update is required.
- Ontology terms: HPO and OMIM terms associated with the gene symbol in VariantGrid. Only displayed when linked term identified.

23.2 Internal gene data

The bottom of the page has 3 grids showing internal data (grids only display when data available)

- Classifications: Summary table of classifications associated with the gene symbol. Click on the links to access the full classification record.
- Variants: A list of all variants located within the genic locus with a Het or Hom_Alt count >= 1 (this excludes low AF somatic variants), as well as any variant that has been tagged or classified in the database (warning: classified/tagged variants may include somatic variants). Columns in the Gene Variants grid below are based on

your User Settings. Change your default column selection to alter the display. To explore the data in the grid click the filter link to display the advanced filter controls.

- Gene Lists: A table of all user entered gene lists containing the gene symbol.

CHAPTER

TWENTYFOUR

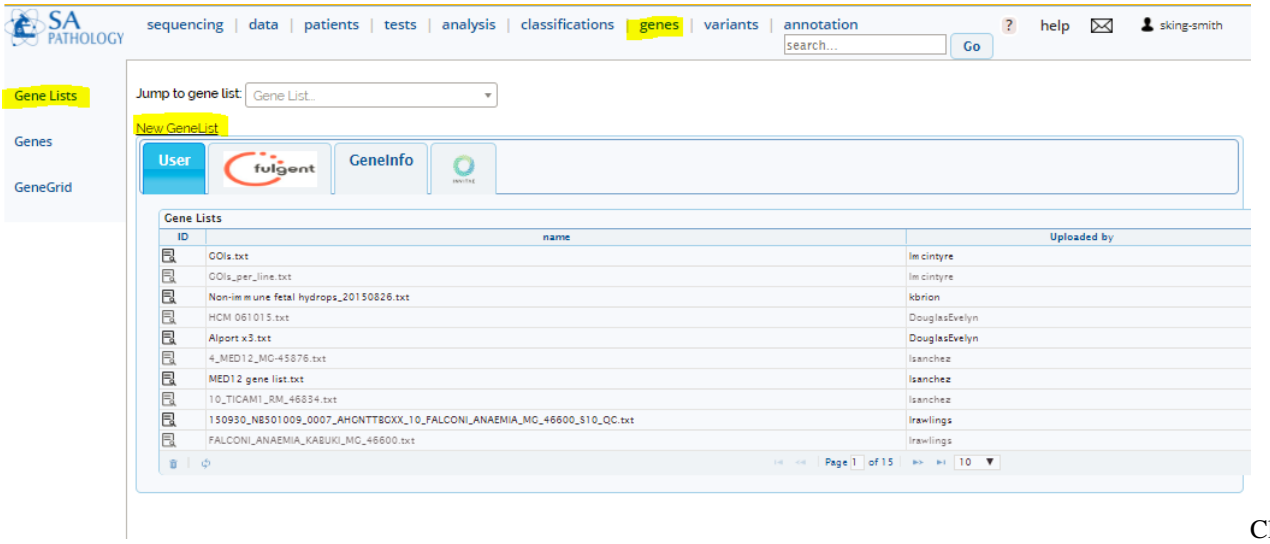
GENE LISTS

Menu: **[genes]**

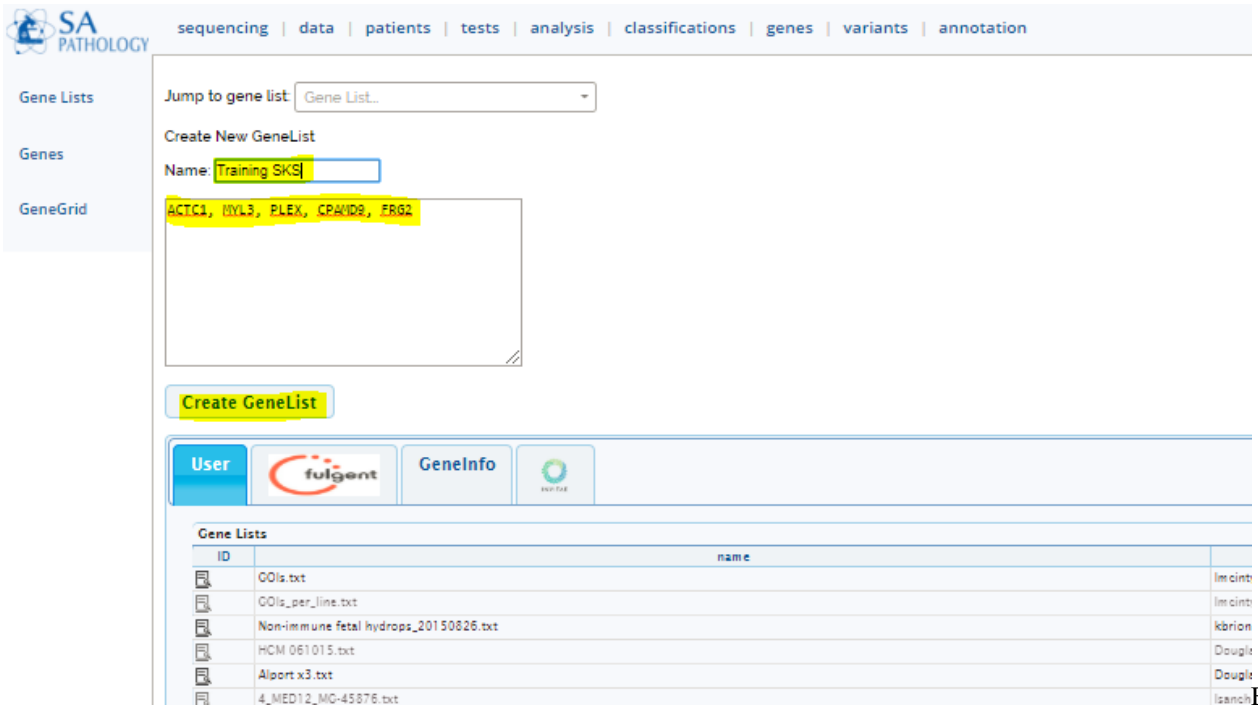
24.1 Creating Gene Lists

Ways to create a gene list include:

- Upload a text file (see *upload*)
- Create via GeneGrid
- Creating manually (see screenshot below)



Click on New GeneList

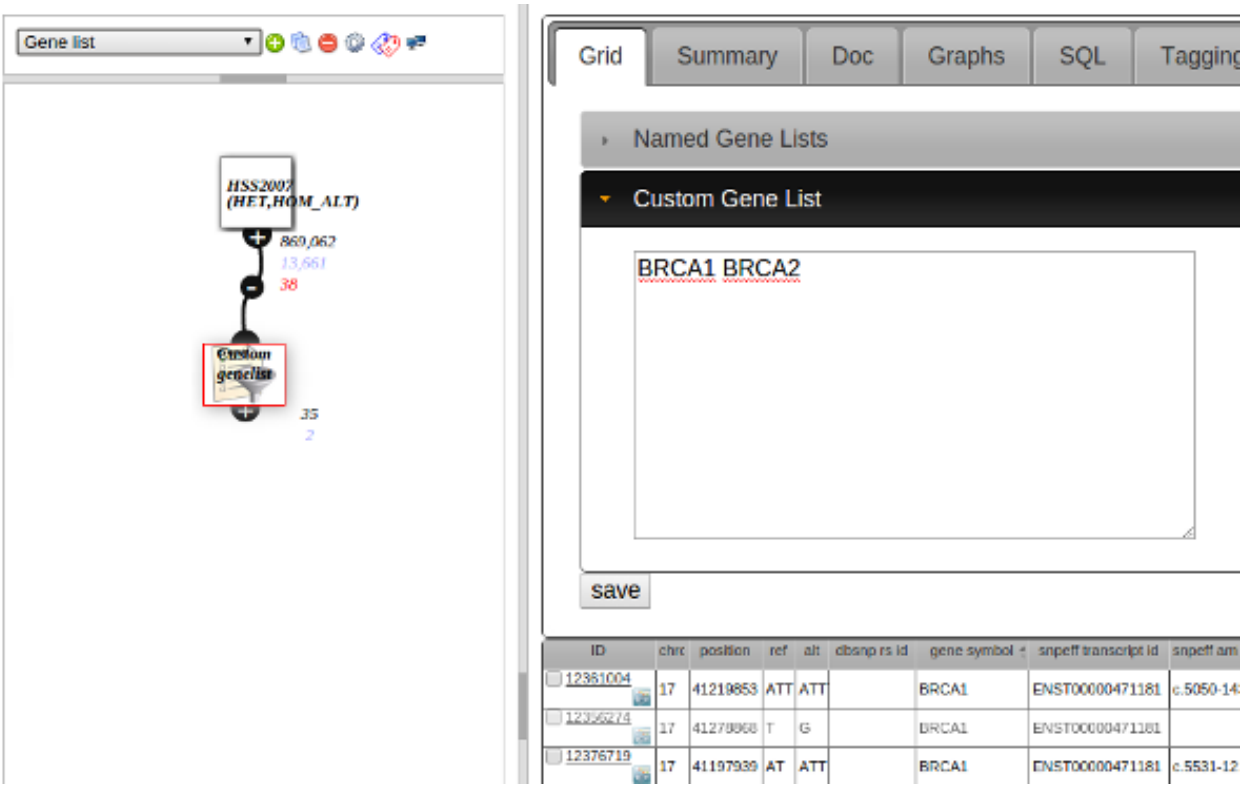


Enter name, genes and click save

24.2 Using gene lists in an analysis

In an *analysis*:

1. Add and connect a GeneList Node
2. In the node editor - select a previously created gene list in **Named Gene Lists** or enter gene symbols directly via **Custom Gene List**
3. Click "Save" to filter to those genes
4. You can see what genes are in the list in the "Genes" tab of the node editor


Gene list
HSS2007 (HET,HOM_ALT)
860,062
13,661
38
Custom gene list
35
2
Grid
Summary
Doc
Graphs
SQL
Named Gene Lists
Custom Gene List
BRCA1 BRCA2
save
ID
chrc
position
ref
alt
dbsnp rs id
gene symbol
snpeff transcript id
snpeff am
12361004
17
41219853
ATT
ATT
BRCA1
ENST00000471181
12356274
17
41278868
T
G
BRCA1
ENST00000471181
12376719
17
41197939
AT
ATT
BRCA1
ENST00000471181

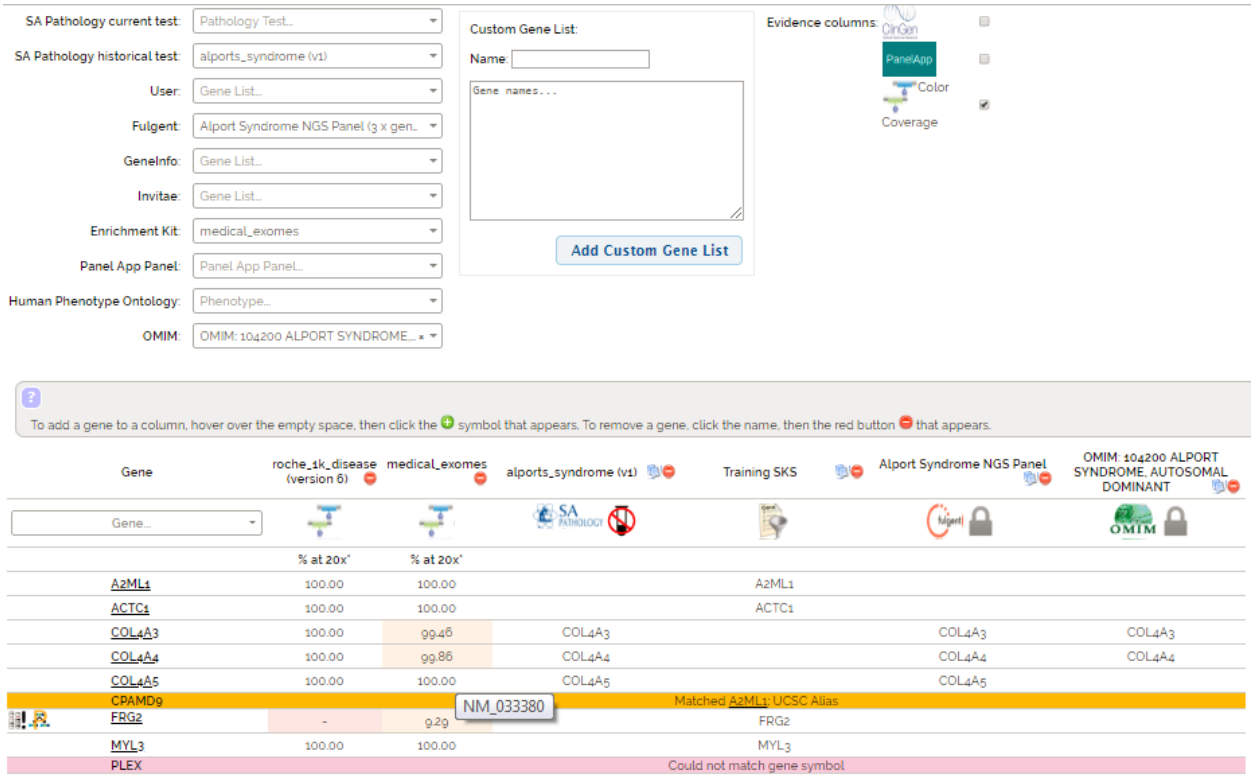
CHAPTER

TWENTYFIVE

GENE GRID

Menu: **[genes] -> [gene grid]**

GeneGrid allows quick comparisons between gene lists and adding/removing genes from them. Genes are rows and gene lists are columns.



25.1 GeneGrid screen

You can copy/paste the URL at any time to re-create a particular comparison.

Choose lists from the top left select boxes, or manually paste in gene names into the **Custom Gene List** text entry box.

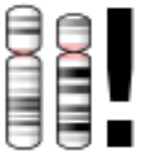
Click the red delete button to remove a gene list column.

In the top right are optional evidence columns which provide information about genes.

See Gene Coverage for details on how the **% at 20x** values in the Enrichment Kit columns are calculated. Enrichment kits are automatically added when a *pathology test* that uses it is added to the grid.

25.2 Gene Info

Small icons next to gene names on the left of the grid indicate the gene has one of these attributes:

 Alternative Haplotype  Pseudogenes  Triplet repeat disorders

CHAPTER

TWENTYSIX

GENE COVERAGE

Gene Coverage refers to how well a gene was covered by high throughput sequencing reads. This is useful to know how confident you can be about a lack of variant calls in a region.

Having gene coverage associated with a VCF sample allows you to be warned in an *analysis* when a gene in a gene list is below a threshold (default: 20x) and you may be missing some variants. The node will flash yellow, and the "genes" tab will be highlighted yellow so you can view which genes have low coverage.

Where gene coverage has been uploaded (eg on diagnostic systems where QC is automatically uploaded) box-plots of sample coverage for a gene will be shown on the gene symbol page

26.1 Canonical Transcripts

Many genes have multiple transcripts, but people want only one value for each gene.

This is achieved by choosing a single (representative or canonical) transcript, and use that transcripts value for the gene.

A CanonicalTranscriptCollection is a list of gene:transcript mappings imported into the system. The administrator can import different collections, linking them to EnrichmentKits and setting a system default.

26.2 Sample QC metrics

You can *upload* gene coverage files (.txt files) which use the system default canonical transcripts. You can then associate them with a *sample from a VCF*

Sample QC coverage loaded via *sequencing features* - and automatically choose transcripts based on EnrichmentKit

26.3 GeneGrid EnrichmentKit coverage

The per-gene QC metrics for an EnrichmentKit on the GeneGrid page are from *Gold Standard Runs*, using the canonical transcripts for that EnrichmentKit.

CHAPTER

TWENTYSEVEN

PATHOLOGY TESTS

Note: This is a diagnostic specific feature which may not be enabled on all systems

Menu: **[tests] -> [manage tests]**

Pathology Tests are curated, versioned gene lists offered as a diagnostic test. There can be multiple versions of a test.

A Pathology Test Version is a specific versions of a pathology test.

27.1 Active tests

Each pathology test has at most one currently active test - the one available for test orders.

An active test is the most recent confirmed version of a pathology test.

Active test logo

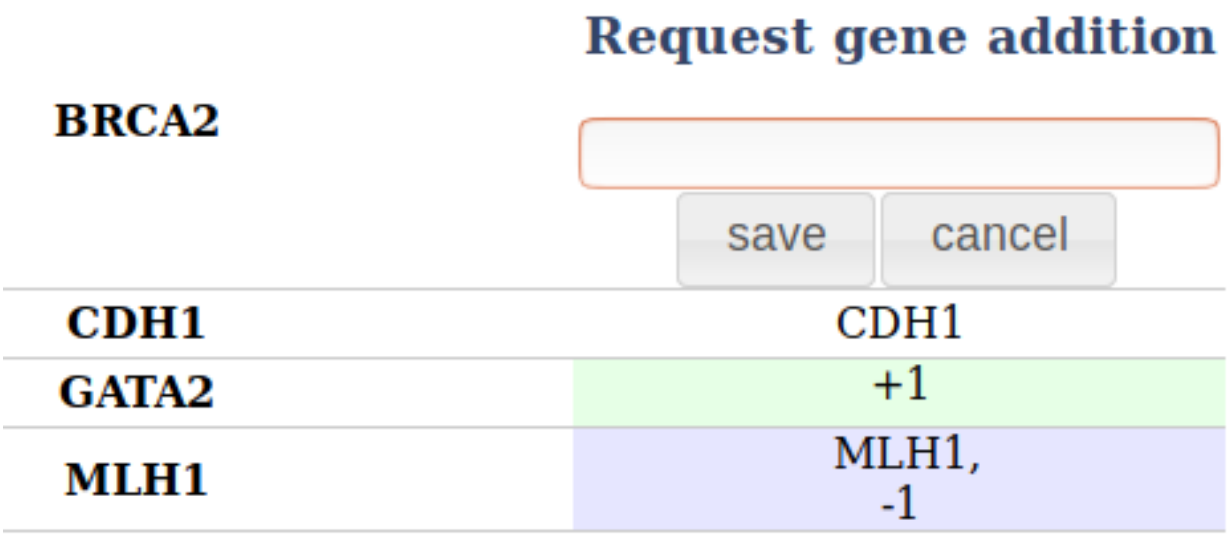
All other versions of tests

The curator confirms & adds a time-stamp by clicking the **Confirm Test** button. Once a test has been confirmed it cannot be modified, and any further changes must create a new test version.

27.2 Requesting gene changes

Only the curator can modify a test, everyone else can make modification request but these must be approved by the curator. Contact an administrator to change curator for a test.

Make gene modification requests on the *GeneGrid* page.



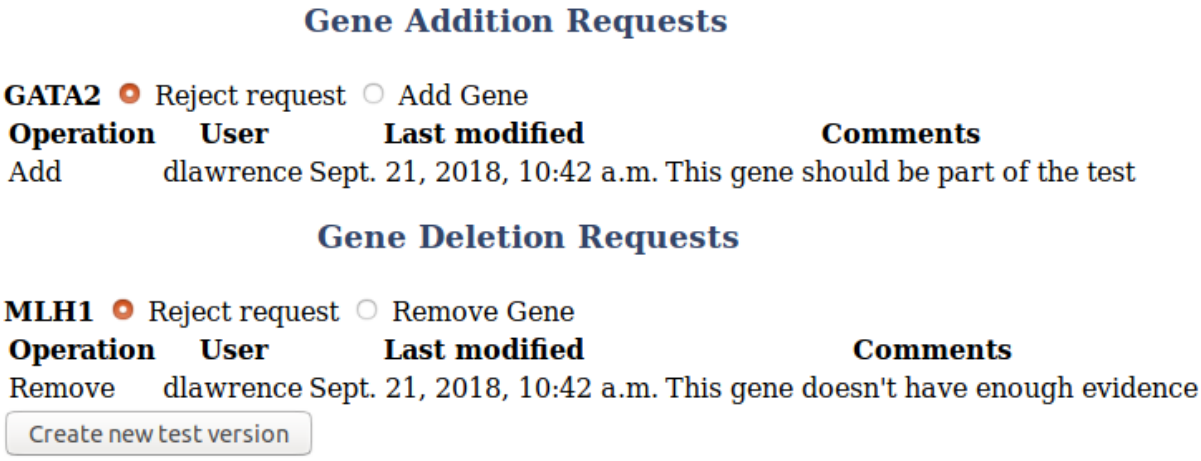
The gene symbols in the pathology test column are always what is in the test. The +/- numbers (green background for add, blue for delete) in the image above are counts of requested additions/removals for that gene.

To request a gene addition: Add genes to the *GeneGrid*, then click on an empty space where the gene should be. *To request a gene deletion:* Click on an existing gene, then the red delete symbol which appears.

In both cases a box will appear where you can enter a brief justification of the request. Only put a brief summary - please put in depth evidence such as linking a disease with a gene or adding literature on the gene page (click on the the gene name on the left column of the grid to open gene page in a new window).

27.3 Accepting gene changes

The curator can see any pending requests on the pathology test version page, where they can accept/reject them.



Any genes added will have the user, date and brief justification comment from the addition request stored on the "Modification info column" which you can see on the grid of genes for a pathology test version.

The outcomes for any processed requests can be seen by all users at the bottom of the page:

Outcome	Operation	User	Last modified	Comments
Accepted	Add	dlawrence	Sept. 21, 2018, 10:44 a.m.	This gene should be part of the test

Outcome	Operation	User	Last modified	Comments
Rejected	Remove	dlawrence	Sept. 21, 2018, 10:44 a.m.	This gene doesn't have enough evidence

CHAPTER

TWENTYEIGHT

TEST ORDERING

CHAPTER

TWENTYNINE

PATIENTS

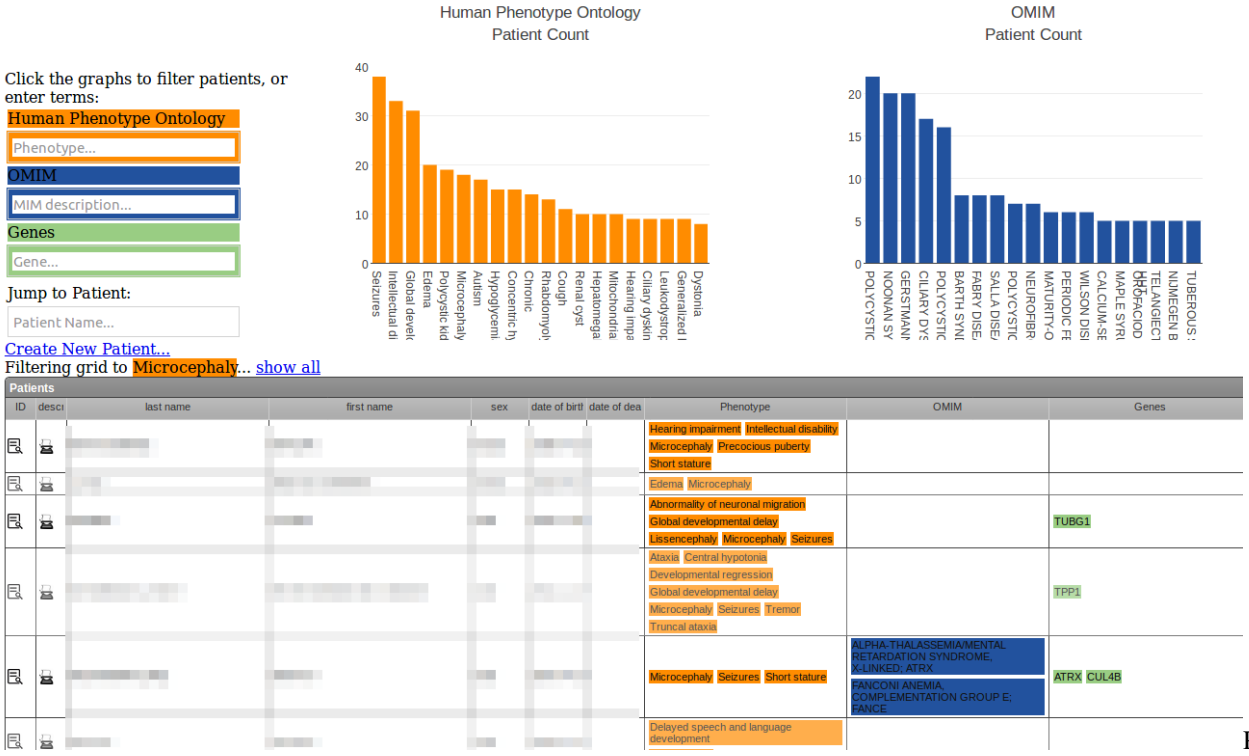
Menu: **[patients]**

Create patients to store phenotype information and link multiple samples (eg tumor/normal) together.

29.1 Searching

You can search by name, code or free text in the phenotype description.

Click the graph of phenotype terms to filter the grid to patients with that phenotype.



Patients grid filtered to microcephaly

29.2 Patient records

Import a CSV to create patients in bulk. Click the **patient record imports** link at the top of the page, then can select to download an example CSV with your samples pre-filled, so it's easy to match your patients to your existing data.

You can also create patients one at time via a form, by clicking the **Create New Patient** link just above the grid.

29.3 Other sources of patients

Patients can be created via the patholoy test ordering system.

On a private server (eg diagnostic lab intranet), patient records can be automatically created via your LIMS/Patient records system (speak to your administrator)

29.4 Other

Family Code is useful for linking together patients

The system can be configured to show/hide names, or convert birthdates to years depending on your privacy needs.

CHAPTER

THIRTY

PHENOTYPES

It is useful to store phenotypes, diseases and genes for a patient. Having this information well structured and using controlled terms is very useful as it allows us to:

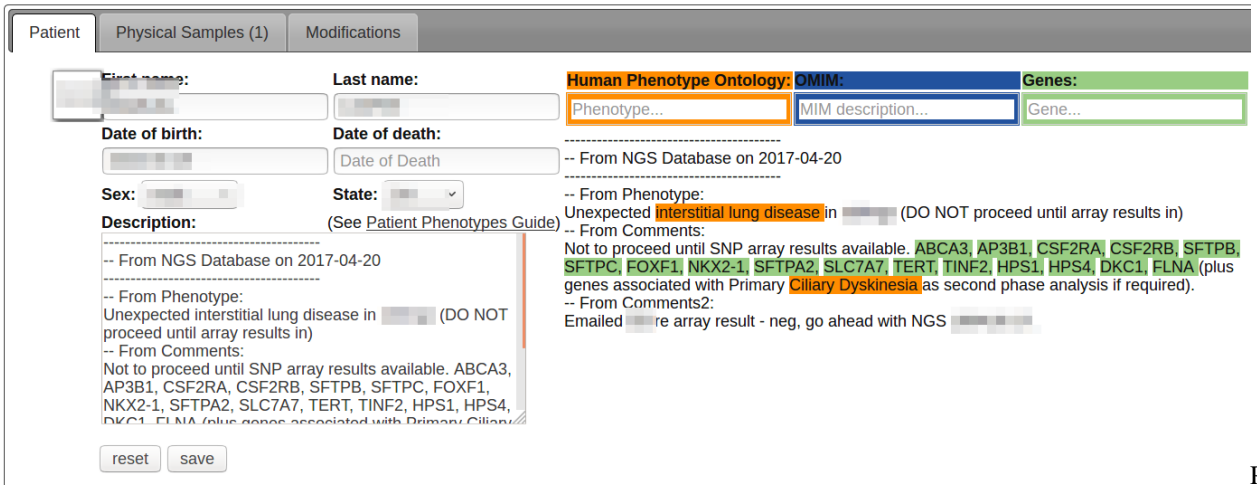
- Filter variants to genes associated with a disorder
- Know phenotypes for patients that share variants
- Perform analyses across disease cohorts (is the same variant or gene responsible for the disease or are they different?)
- Track per-disease solve rates

30.1 Assigning Terms to Patients

You can auto-complete terms in the boxes, which will be added to the bottom of the patient description.

Or, you can type plain text and we'll automatically match your words to Human Phenotype Ontology, OMIM and Gene Names.

Matched terms will be highlighted to the right of the description box.



Patients grid filtered to microcephaly

30.2 How phenotype term matching works

Everything after "–" on a line is ignored and can be used for comments.

The text is broken up into sentences based on punctuation and new lines.

The sentence is separated into words, and then sub sets of the words in order are created, and sorted largest to smallest. For instance:

```
The cat sat on the mat
cat sat on the mat
The cat sat on the
sat on the mat
cat sat on the
The cat sat on
The cat sat
on the mat
sat on the
cat sat on
the mat
cat sat
The cat
on the
sat on
mat
the
sat
cat
The
on
```

This allows us to find the biggest matches first. If a match occurs, the unmatched parts of the sentence continue to be searched until there is nothing left. If no match occurs for a sentence, we try the next smaller one.

Some filtering is done to avoid matching to common words and terms. For instance "Trio" is a gene name, but we will not match it as a gene if the sentence also contains the name of a enrichment_kit or one of the words: "exome", "WES", "father" or "mother".

Matching occurs first against Human Phenotype Ontology terms and synonyms, and OMIM terms and aliases.

If no exact match is found, we try again using mismatches - 1 mismatch (including insertions/deletions) is allowed for two or more words.

For single words, we only allow mismatches if the word is more than 5 letters long and made entirely of letters (ie no digits or symbols).

Single words are then matched (exact with no mismatches) to gene names.

Sometimes there will be multiple matches, eg "PKD1" will map to both the OMIM term PKD1 (POLYCYSTIC KIDNEY DISEASE 1) and the gene PKD1. This is usually what people want as the gene is associated with the disorder.

CHAPTER

THIRTYONE

COHORTS

Menu: **[patients] -> [cohorts]**

A cohort is a collection of samples, which you can analyse as a group. A multi-sample VCF automatically becomes a cohort, but you can create your own to organise your own samples.

31.1 Create a new cohort

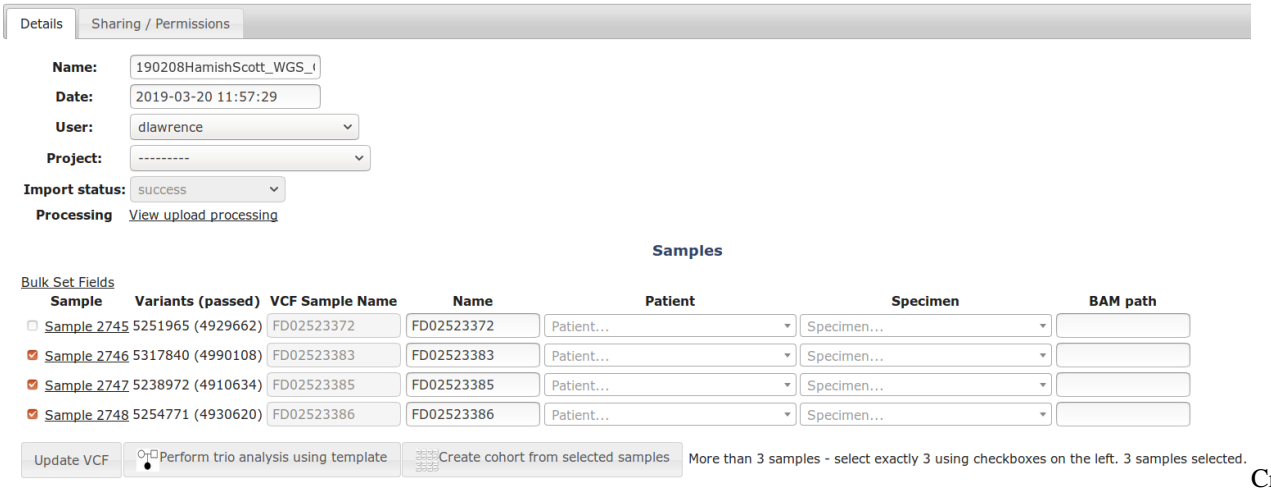
From the cohort page, enter the name of a cohort and click the **Create** button.

This opens the Add/Remove samples tab. Add samples to your cohort by auto-completing sample names in the Enter to add box, or filter the grid, select the checkbox to the left of a sample, and click the green arrow to add, or red button to delete.

Once you have finished adding/removing samples, click save. This processes the cohort so it can be used in analyses.

31.2 Create from a larger cohort

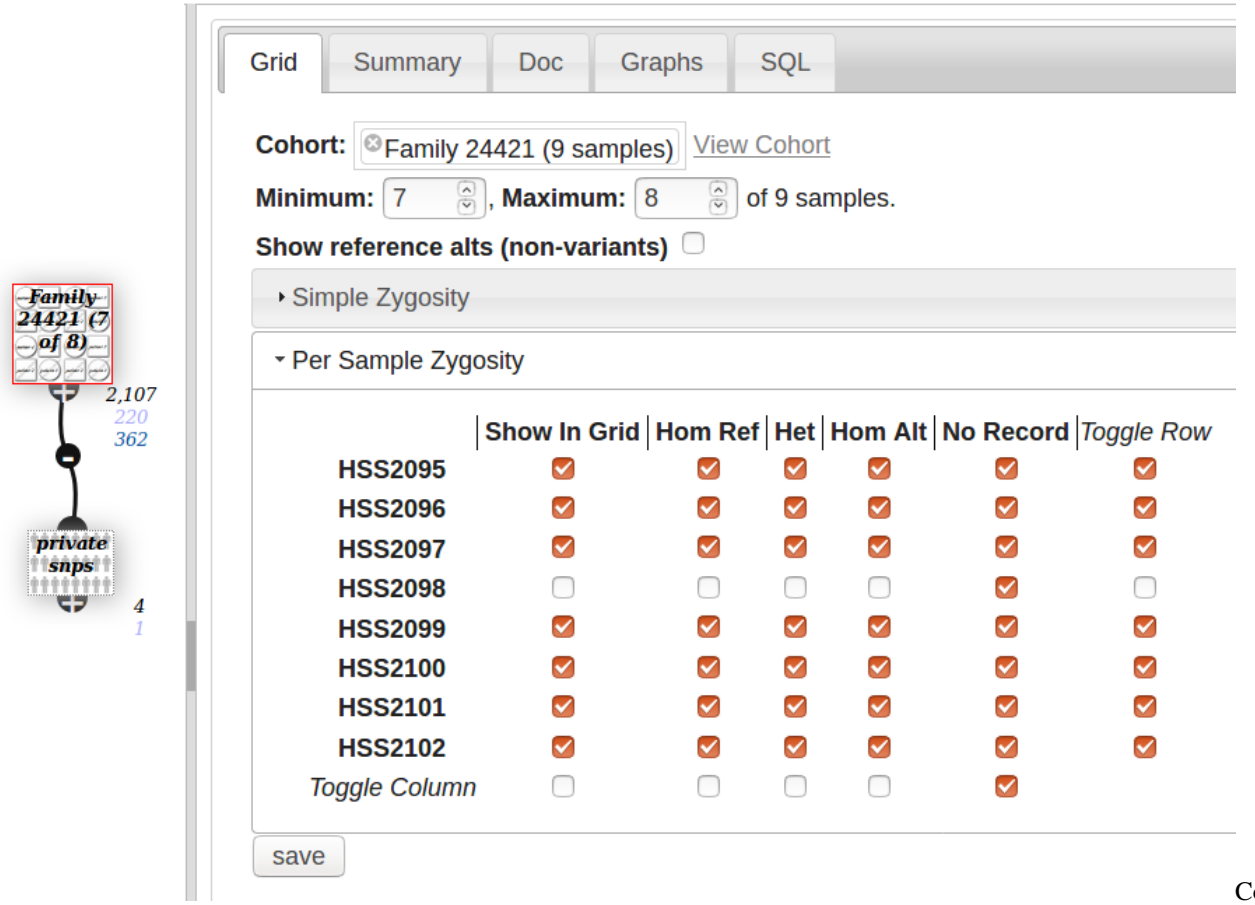
You can create a smaller cohort from a larger one. Select at least 2 samples then click the **[Create cohort from selected samples]** button. Selecting exactly 3 samples allows you to create a Trio which allows for simpler analyses.



Creating a sub-cohort

31.3 Cohort Analyses

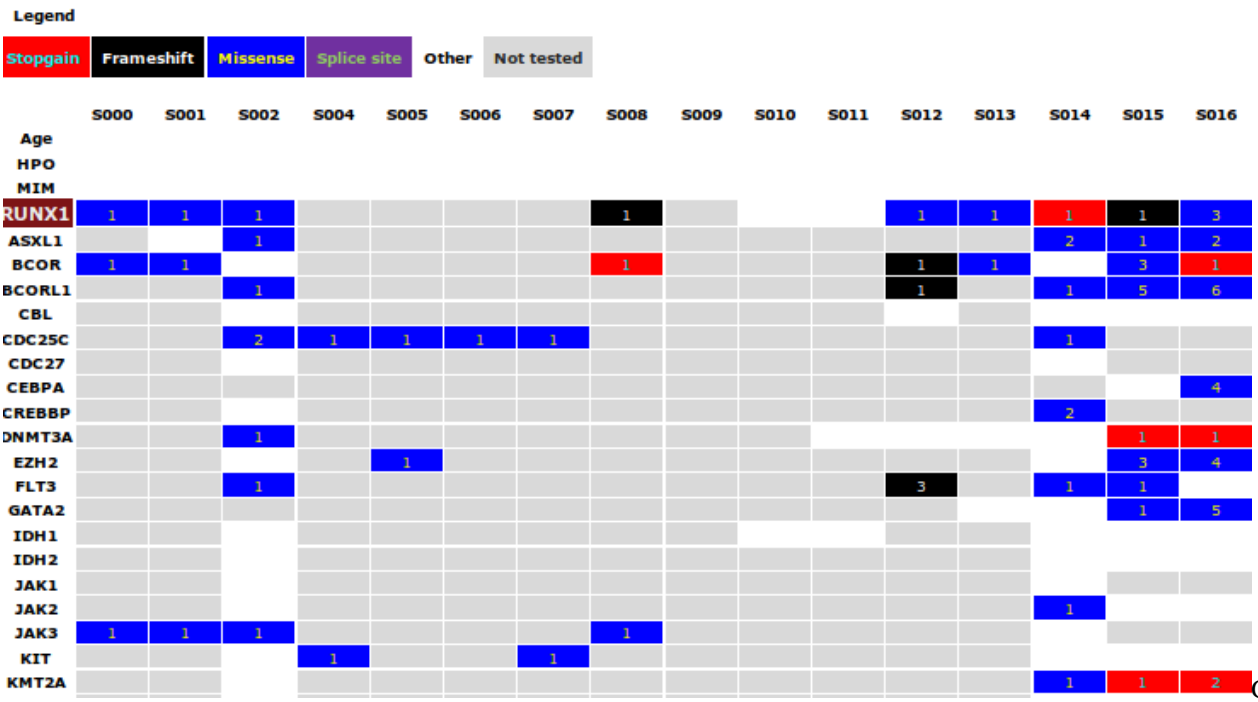
Use the Cohort Node to filter by counts within the cohort (eg in 7 out of 8 of the samples) or zygosity. (see screenshot below).



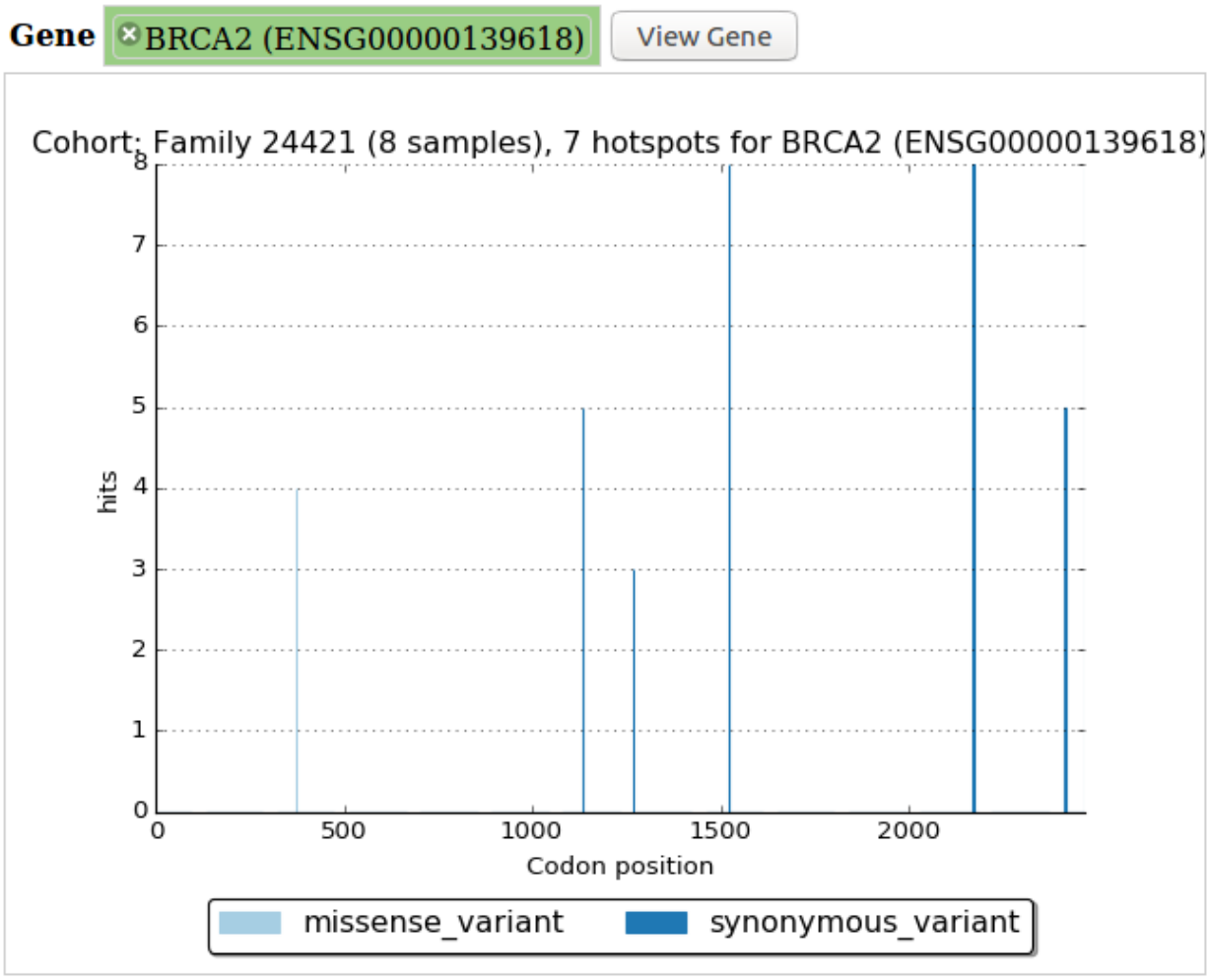
Cohort Node filtering by zygosity

Quickly create an analysis using the cohort by clicking "Create new analysis for cohort" on the details tab of the cohort page.

There are some other analyses you can perform by clicking links at the bottom "Analysis" section of the cohort/VCF page, eg:



Gene/Sample Matrix - Shows number of variants that meet a certain criteria per gene. Access by clicking "View gene damage counts for this cohort"



Cohort

Hotspots graph - access by clicking “View gene hotspots for this cohort”

CHAPTER

THIRTYTWO

TRIOS

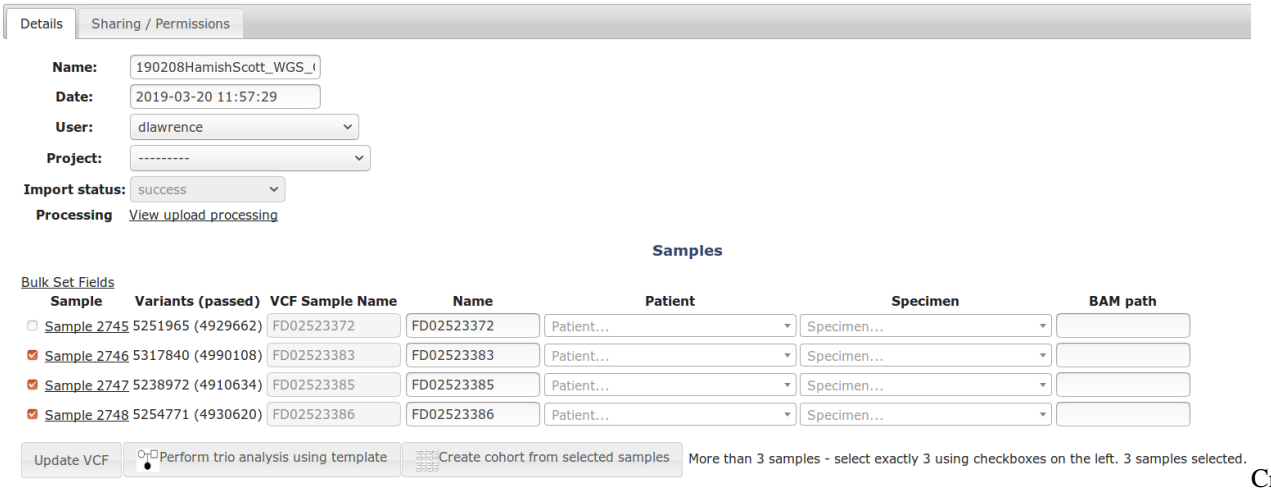
Menu: **[patient] -> [trios]**

A trio is a collection of 3 samples (mother/father/proband) which are frequently analysed together in high throughput sequencing, as they have a number of standard analyses.

32.1 Creating a trio

It is far better to upload a trio within the same *multi-sample VCF*. If not, you must first create a cohort containing the 3 samples/

View the VCF or cohort, select exactly 3 samples then click the **[Perform Trio Analysis using template]** button.



Creating a Trio

The Trio wizard will now open, showing the 3 samples and patient / phenotype info. Assign samples (1 each to mother/father/proband) and check mother or father affected if they also have the disorder.

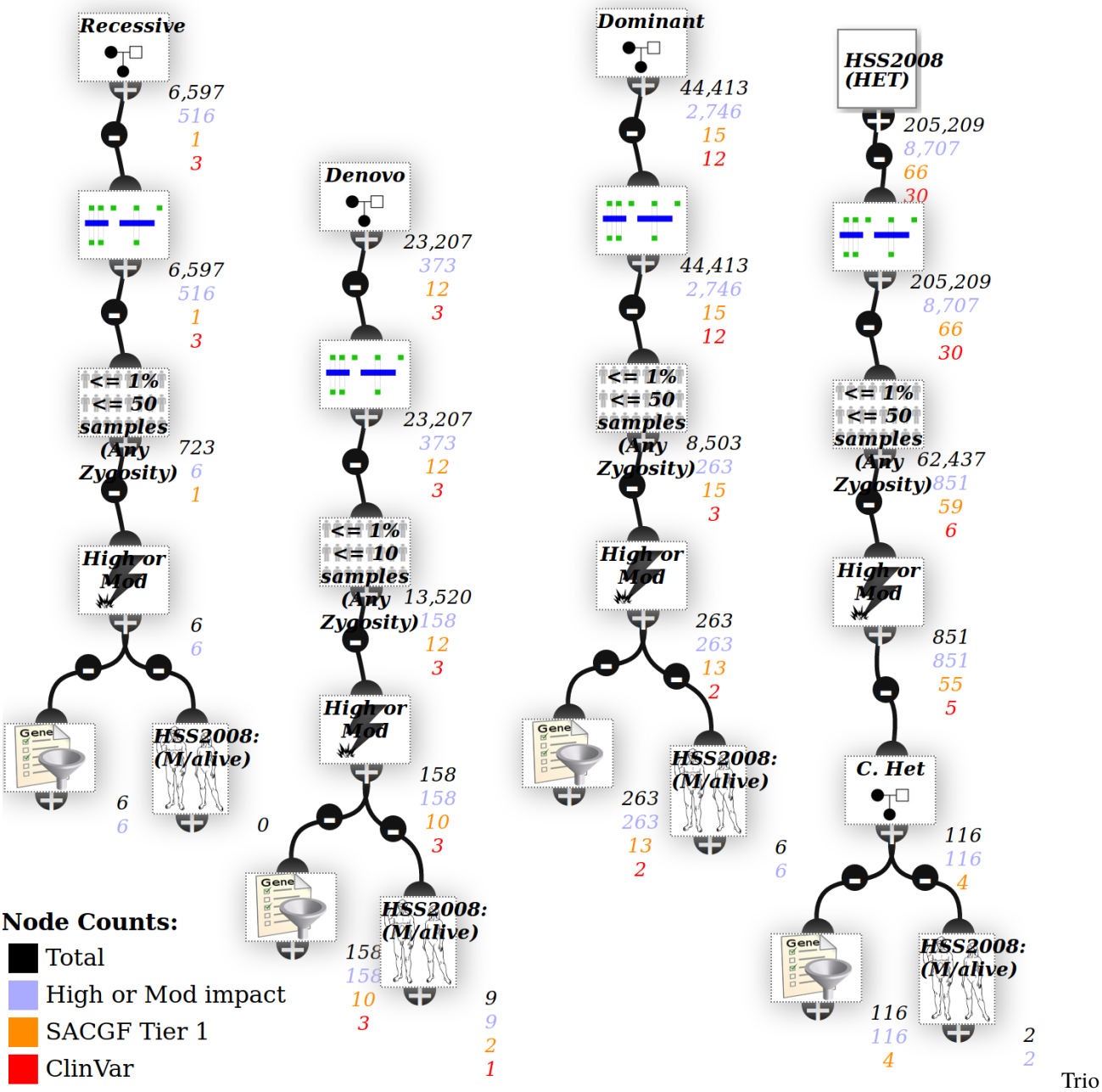
32.2 Digital karyomapping

By checking a trio's zygosity, it's possible to perform a number of relatedness calculations, see *karyomapping*.

A genome-wide count is automatically performed, and a summary provided on the trio page - this is useful for checking for sample mix-ups.

32.3 Trio inheritance analysis

An analysis is created using different inheritance models (see below). If either parent is affected it will also use an autosomal dominant inheritance model.



Trio inheritance analysis

The phenotype at the bottom uses the proband patient phenotypes, and sample gene lists.

32.3.1 Require Zygosity Calls

By default, the filters are strict and require zygosity calls in all patients - for instance the recessive inheritance model requires a variant to be HOM in proband and HET in both parents.

However that may be overly strict - one parent may have low coverage, with no variants recorded at that locus.

Click on an Trio node to open the editor - unchecking the **require zygosity calls** box is less strict and allow for variants that are missing due to low coverage.

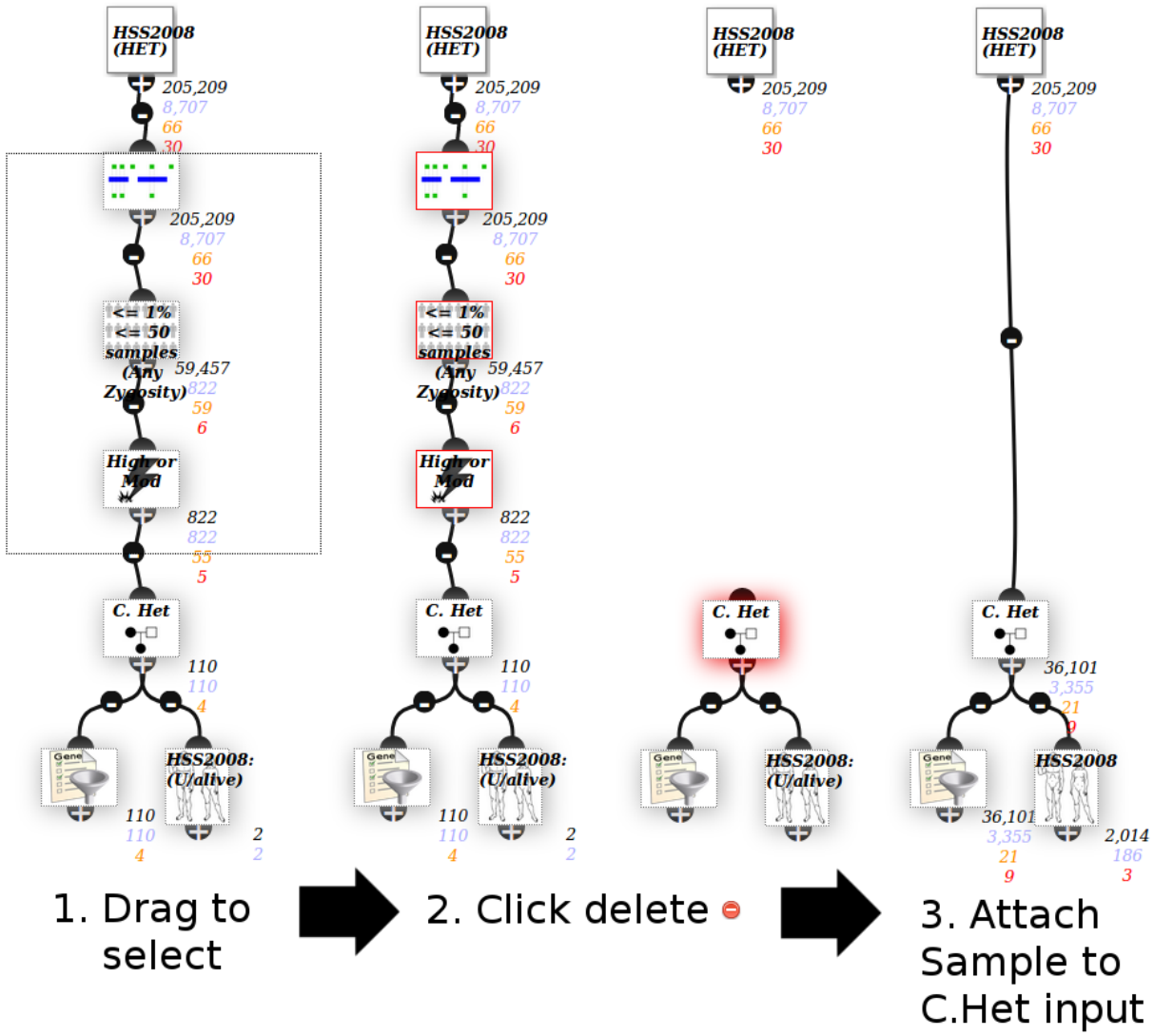
32.3.2 Compound Het filter

Compount heterozygous means 2 variants in the same gene from different parents.

The C. Het node in the bottom right of the screenshot above is a filter node - ie it has another node connected to the top, while the other inheritance models do not.

This is because you probably don't want every gene with >=2 variants, but rather only >=2 damaging/rare ones. Adjust the filters above the C.Het node to adjust this.

Modify the analysis as per instructions below to filter to all of them.


HSS2008 (HET)
205,209
8,707
66
30
205,209
8,707
66
30
<= 1% <= 50 samples (Any Zygosity)
59,457
822
59
6
High or Mod
822
822
55
5
C. Het
110
110
4
Gene
HSS2008: (U/alive)
110
110
4
2
2
1. Drag to select
2. Click delete
3. Attach Sample to C.Het input
36,101
3,355
21
9
HSS2008
2,014
186
3

CHAPTER

THIRTYTHREE

PEDIGREE

Menu: **[patient] -> [pedigree]**

Pedigrees describe family relationships, and marks samples as "affected/unaffected"

This and can be used to filter for inheritance models (eg recessive/dominant) in an analysis.

For the common case of 3 samples, perhaps use a Trio

To create a pedigree

- Create a .ped file for your family, eg using the Phenotipes editor
- Upload the .ped file to VariantGrid
- Match the ped file family and cohort samples to create a Pedigree
- Use an analysis *PedigreeNode* to apply inheritance models

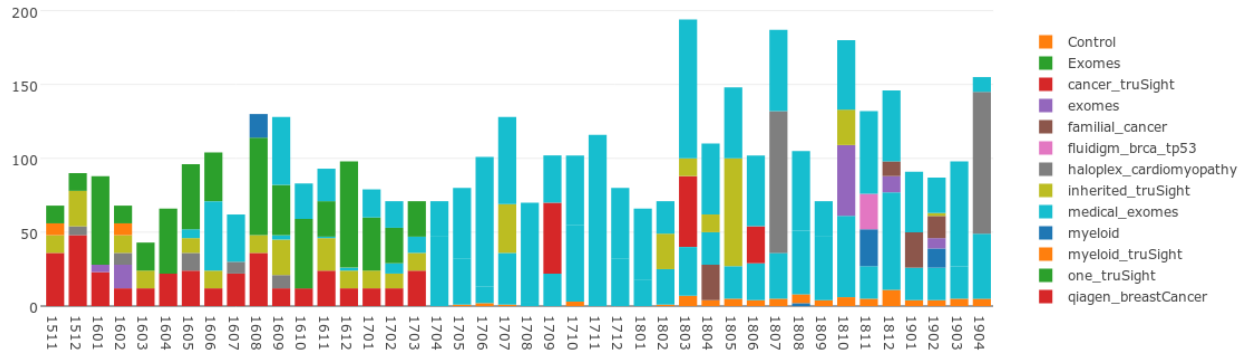
CHAPTER

THIRTYFOUR

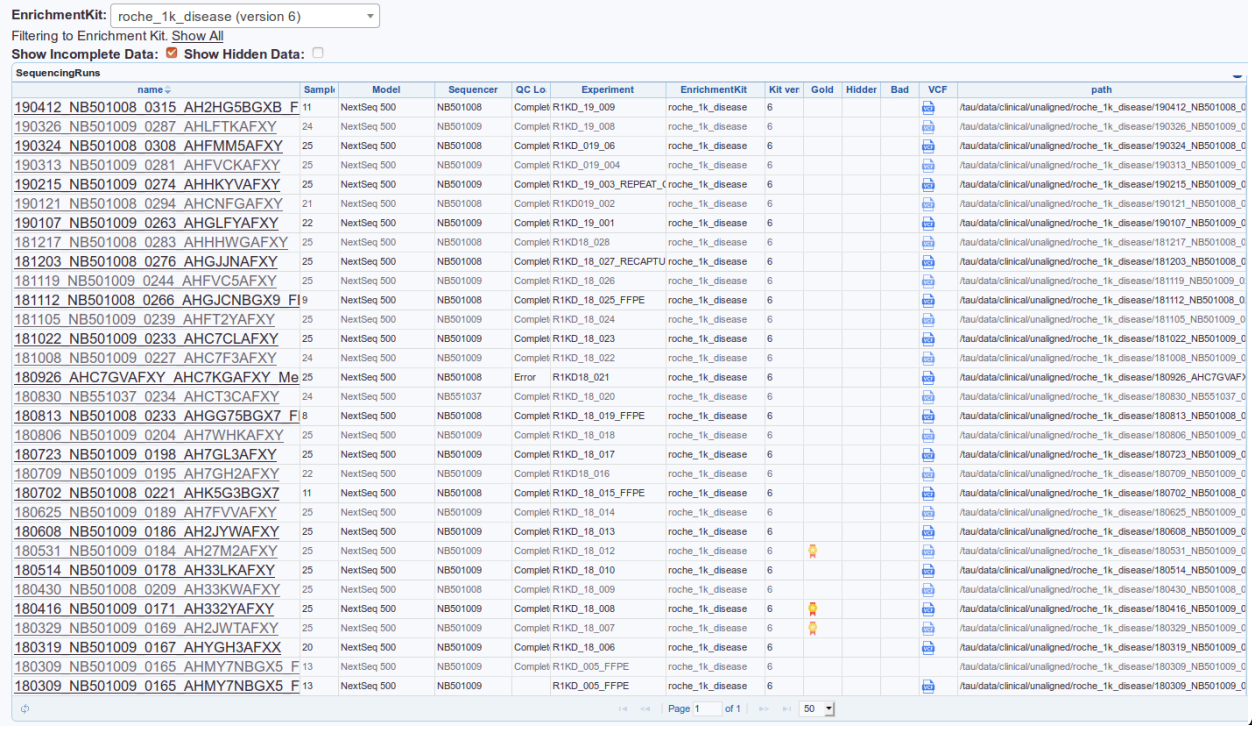
SEQUENCING RUNS

Note: This feature may not be enabled on all systems as it requires access to a network drive (eg a diagnostic lab intranet)

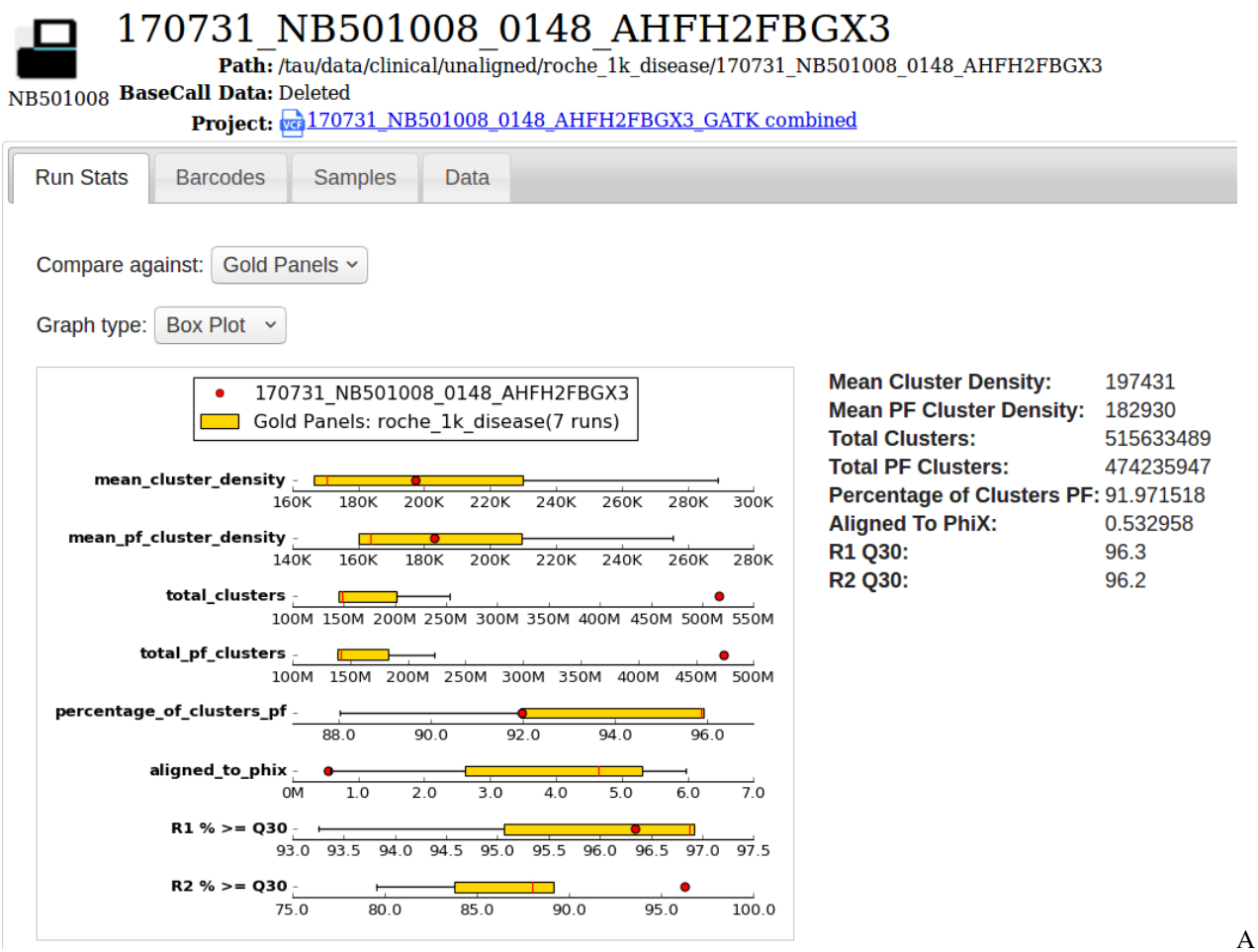
VariantGrid can be setup to automatically scan network disks for sequencing runs to collect QC metrics, gene coverage and automatically load VCFs.



Sequencing Samples over time



Automatically loaded sequencing runs + VCFs



A Sequencing Run

We collect Sequencing QC metrics and display them with interactive graphs. Collecting data over time allows us to see how this run compares to other runs over time (or vs *gold standard runs*).

34.1 Enrichment Kit

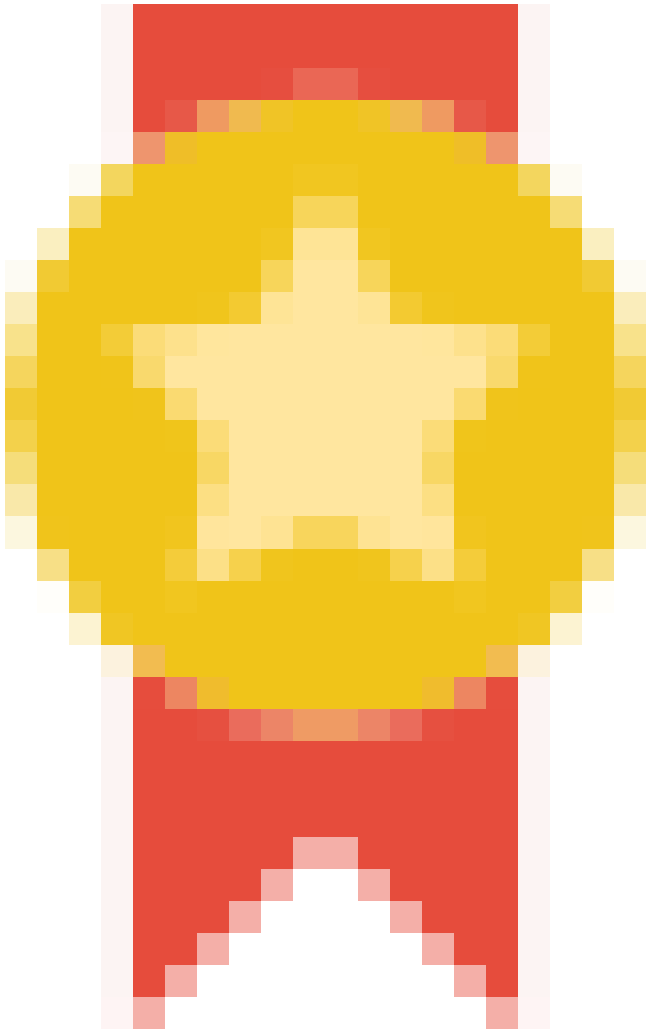
An EnrichmentKit is a lab method to enrich a sample for the DNA regions you are interested in. For instance an exome or custom gene capture kit, or amplicons.

You can set a bed file, a gene list and canonical transcript collection

See VariantGrid Admin docs for more information.

34.2 Gold Standard Runs

The administrator can mark a sequencing run as "Gold Standard" - which means it has passed validation / is of sufficient quality to be used as a benchmark for other runs.



Gold standard runs have an icon ( ) on the sequencing run grid.

Gold runs for an enrichment kit are used:

- In boxplots on QC metrics pages for a sequencing run or other sample QC graphs.
- To calculate average *gene coverage* on the *GeneGrid* page.

34.3 Finding sequencing data

Sequencing Runs are found by searching for the file 'RTAComplete.txt' on the server disks. You can ignore flow cells by putting a file ".variantgrid_skip_flowcell" in the directory.

34.4 Triggering a manual scan

Administrators, or users who have been give the permission "SeqAuto scan initiate" can

Menu: **[sequencing]**, then **manage disk scans** link, then click the button **Scan Disk for Sequencing Data**

CHAPTER

THIRTYFIVE

USER DETAILS

At the top of the page you can set your avatar image, and change your name/email etc.

The avatar is only used on the labs page, **[Classification] -> [Labs]**

35.1 Groups

Groups are used to share data (analyses, classifications etc) between users. If you search for a user in the search bar, you can see groups you have in common with them (so can use to share things by assigning permissions on objects for that group)

Your groups are set by administrators.

There are two groups that every user is a member of:

- Logged_in_users - visible to anyone with a login
- Public - visible to everyone (if in the future we allow access w/o a password)

35.2 Initial group permissions

This lists your groups, whether they are associated with a lab or not. Labs are used for classification share levels.

The check boxes can be used to set initial object permissions, for instance if you set “read” for “mylab” then every time you uploaded a VCF, or create an analysis, it would be visible to people in “mylab”.

This is just the initial setting, you can always click the “sharing/permissions” tab on an object then modify it later.

35.3 Node counts

These are how the node counts are set when an analysis is created. You can always adjust each analysis node counts via analysis settings.

35.4 User Settings

There are multiple levels of settings:

- Global (set by administrators per server)
- Default Lab's Organisation
- Default Lab
- User

The later settings can be used to overwrite the earlier ones if they don't like the defaults.

- **Email Regular Updates** - Opt into email list (Only used for Shariant)
- **Columns** - Default columns for analysis grids (can be changed per analysis)
- **Default Sort by Column** - Default value to sort analysis grids (can be changed per analysis)
- **Tag colors** - Set of colors assigned to tags (modify/create these in 'Tag settings')
- **Variant Link in Analysis Opens New Tab** - Whether left click by default opens up variant details in new tab. No is to open details in the node editor location. It's always possible to right click and select 'open in new tab'
- **Tooltips** - Show/hide help popups on mouse hover
- **Node Debug Tab** - If true, an extra tab appears in analysis node editor, with details about node settings + SQL query.
- **Import Messages** - Get internal notification (message icon top right) when imports are done (eg VCF finished processing and annotating)
- **IGV Port** - Port to connect to IGV on your machine, see IGV Integration
- **Default Genome Build** - Used for search (jump to result if that is the only one for this build) and populating defaults everywhere
- **TimeZone (for downloads)** - Time/date used in classification download
- **Default Lab** - Lab used for creating classifications (you can belong to more than 1 lab)

CHAPTER

THIRTYSIX

CUSTOMISE COLUMNS

You can customise the columns that appear in an analysis grid.

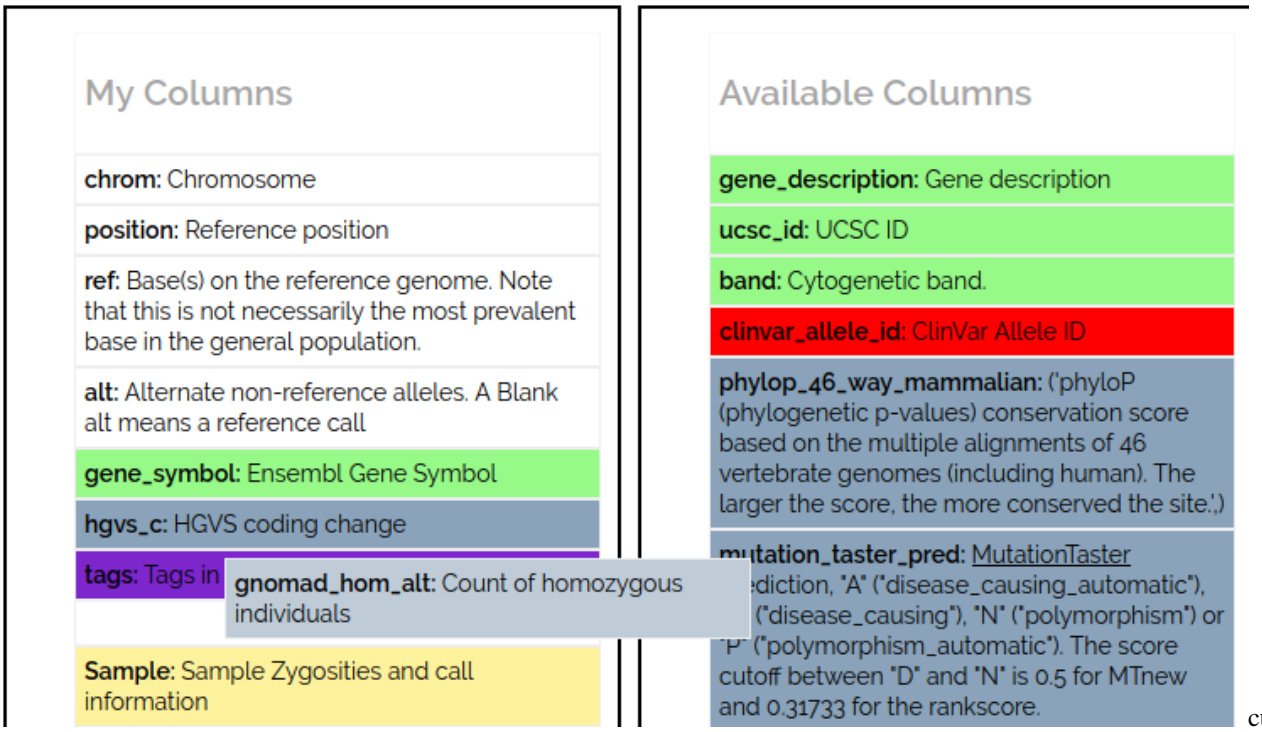
To create a new set of columns, visit the **Customise Columns** (**[user name]->[customise columns]**) page.

You can't modify built-in custom columns, as they are shared by everyone. Click the **[Clone...]** link on the custom columns grid to make a copy and edit it.

36.1 Changing columns

The customise columns page shows grid columns as blocks, which you can drag & drop to add and remove, or change order.

Columns in "My Columns" are in this set, while unused columns are in "Available Columns". The screenshot below shows the user adding the "gnomad hom alt" column after "tags".

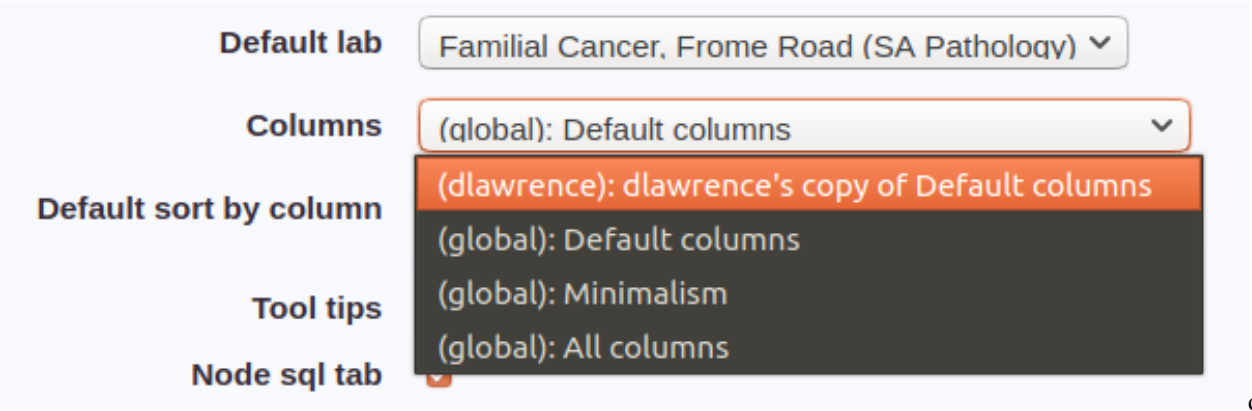


custom columns screenshot

The order of columns (top to bottom) determines the left to right order in the grid.

36.2 Default columns

New analyses are created with columns set to your default columns, which you can change on the user settings (click on **[user name]**)



default columns

36.3 Changing columns in an analysis

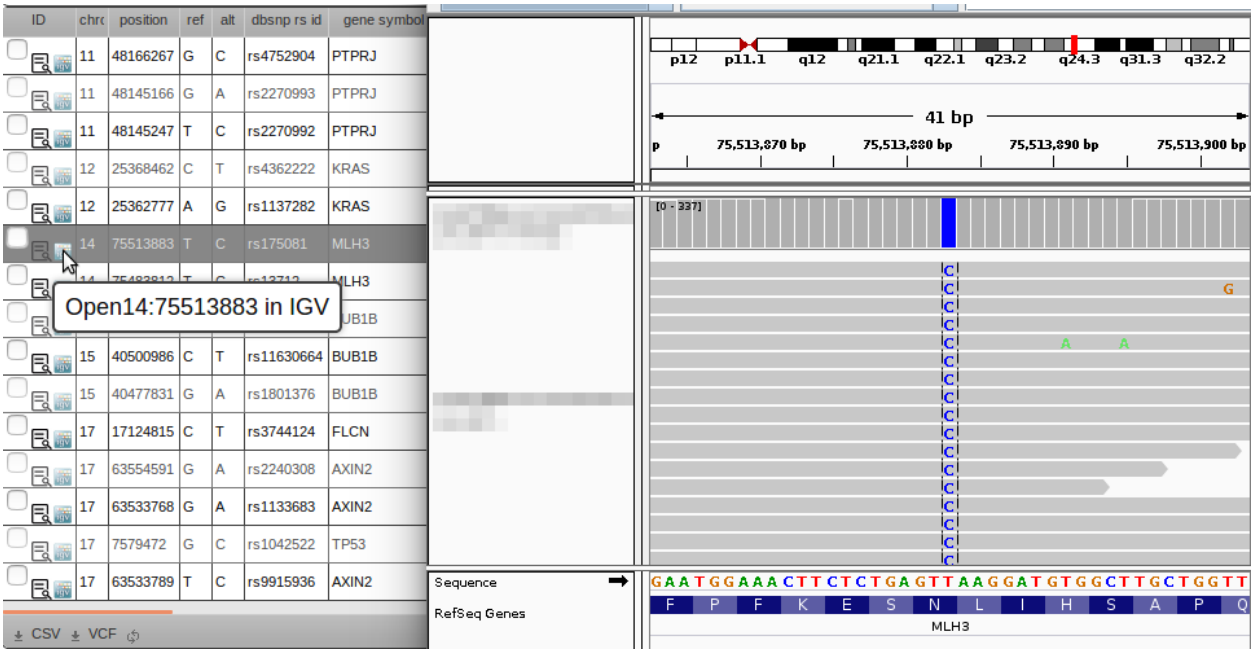
In an analysis click the Settings icon to open the *analysis settings* page, where you can select columns from a dropdown.

CHAPTER

THIRTYSEVEN

IGV INTEGRATION

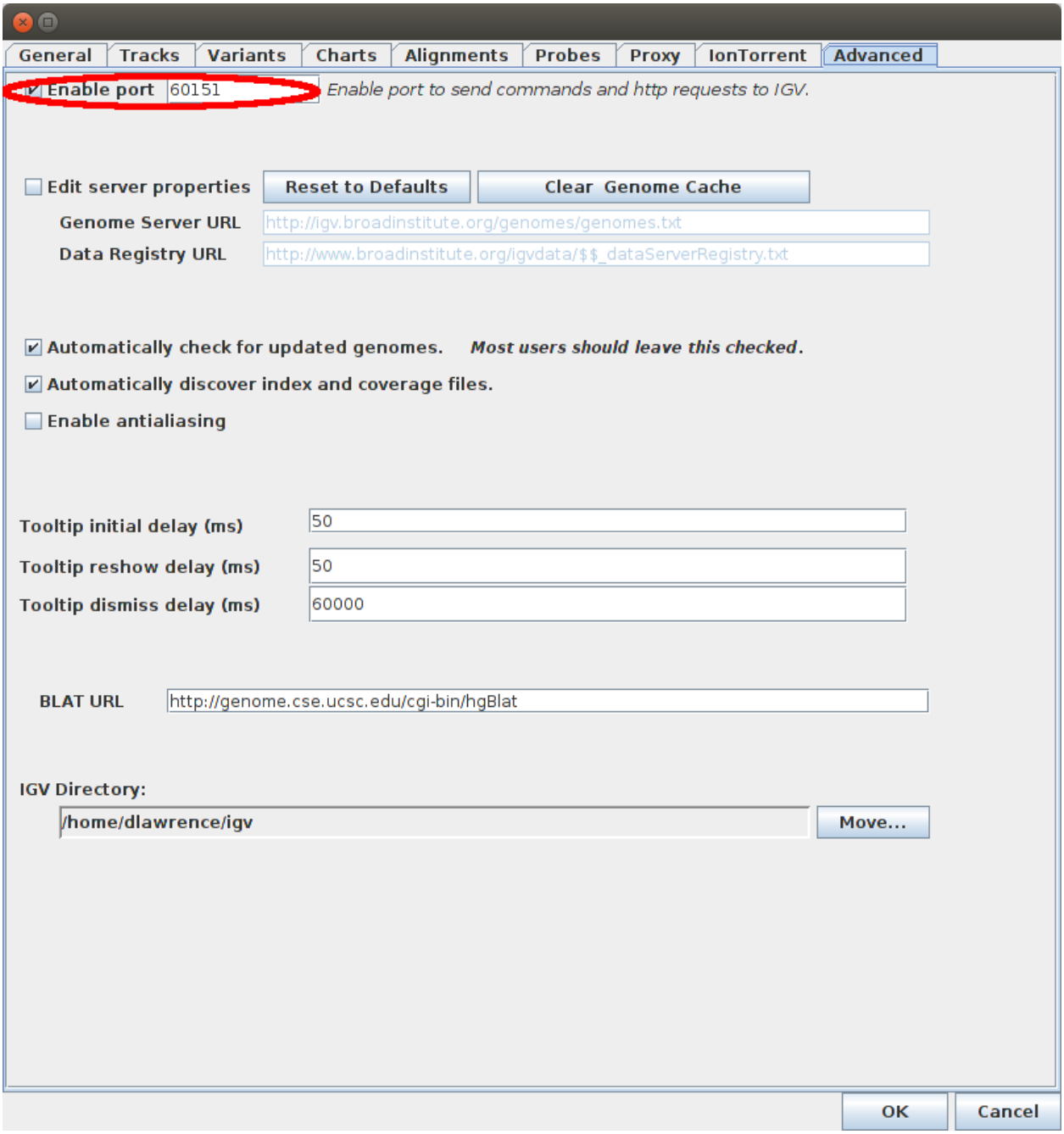
Click the  IGV link to automatically jump to your variants + BAM files in IGV.



37.1 IGV Configuration

IGV needs to be running, and have the Enable Port option ticked.

To check this open preferences in the IGV menu: **[View] -> [Preferences] -> [Advanced] Tab**.


General
Tracks
Variants
Charts
Alignments
Probes
Proxy
IonTorrent
Advanced
Enable port
60151
Enable port to send commands and http requests to IGV.
Edit server properties
Reset to Defaults
Clear Genome Cache
Genome Server URL
http://igv.broadinstitute.org/genomes/genomes.txt
Data Registry URL
http://www.broadinstitute.org/igvdata/$$_dataServerRegistry.txt
Automatically check for updated genomes.
Most users should leave this checked.
Automatically discover index and coverage files.
Enable antialiasing
Tooltip initial delay (ms)
50
Tooltip reshow delay (ms)
50
Tooltip dismiss delay (ms)
60000
BLAT URL
http://genome.cse.ucsc.edu/cgi-bin/hgBlat
IGV Directory:
/home/dlawrence/igv
Move...
OK
Cancel

37.2 VariantGrid Configuration

If the value of the IGV port is different from **60151** (default), you need to change the IGV Port option in your User Settings page.

Clicking the IGV link ( IGV link) will jump to the locus, and show BAM files associated with input samples (Sample or Cohort ancestors). These are the same samples that have their zygosities/allele depth shown on the grid.

Each sample has a bam file path entry. If your samples were automatically loaded from a server, this is probably already set. Otherwise you can change it on the Sample or VCF (VCF) page.

You can set all the samples in a VCF file at once in the vcf page, click Bulk Set Fields to set all samples according to a pattern based on the sample name.

Samples

Bulk Set Fields

BAM path /data/{name}.bam Set Bam Path

Public Data Toggle

Sample	Variants (passed)	Name	Patient	Physical Sample	BAM path	Public Data
	12607 (12264)			---------	/tau/data/clinical/align	
	12512 (12163)			---------	/tau/data/clinical/align	
	12590 (12249)			---------	/tau/data/clinical/align	
	12762 (12420)			---------	/tau/data/clinical/align	
	12768 (12417)			---------	/tau/data/clinical/align	
	12905 (12549)			---------	/tau/data/clinical/align	
	12702 (12357)			---------	/tau/data/clinical/align	
	12770 (12423)			---------	/tau/data/clinical/align	
	12579 (12229)			---------	/tau/data/clinical/align	
	12643 (12297)			---------	/tau/data/clinical/align	
	12585 (12247)			---------	/tau/data/clinical/align	

37.3 Network drives and File Servers

Many labs access data via servers, or network shares. These can be different on different computers.

It is recommended that you set bam file path to be the location on the server, so that it is consistent between users.

Different data access methods on different computers can then be managed by having users change their configuration on the IGV Integration page.

CHAPTER

THIRTYEIGHT

CLASSIFICATIONS

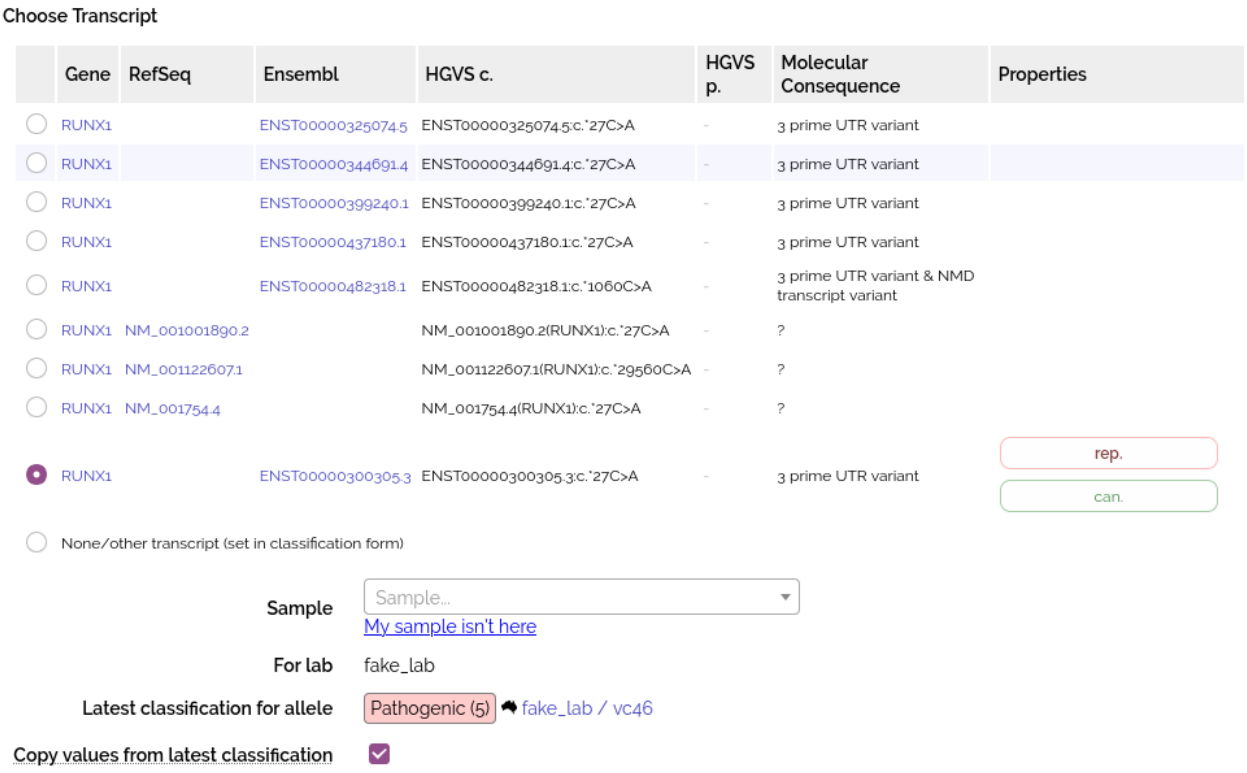
38.1 Creating Classifications

Create classifications as follows:

- From an analysis (see *analysis classification workflow*)
- From an existing variant via the *variant details page*
- Via API (See Shariant API docs)
- Entering a HGVS name into the box at the top of the classifications page.

38.2 Create from existing variant

When you click "New Classification" from the allele or variant details page, you are shown a form to pick the transcript and sample:



A number of fields are auto-populated from *annotation* and sample information (data from VCF record, patient phenotype etc).

Classifications made against a sample are linked from the bottom of the VCF and sample pages.

Variants created from the external API are not auto-populated from annotation.

38.3 Editing

See the Classification Form.

38.4 Configuring Fields

An administrator can add/remove EvidenceKeys which are used to create fields.

They can also hide visible fields on a per-lab basis.

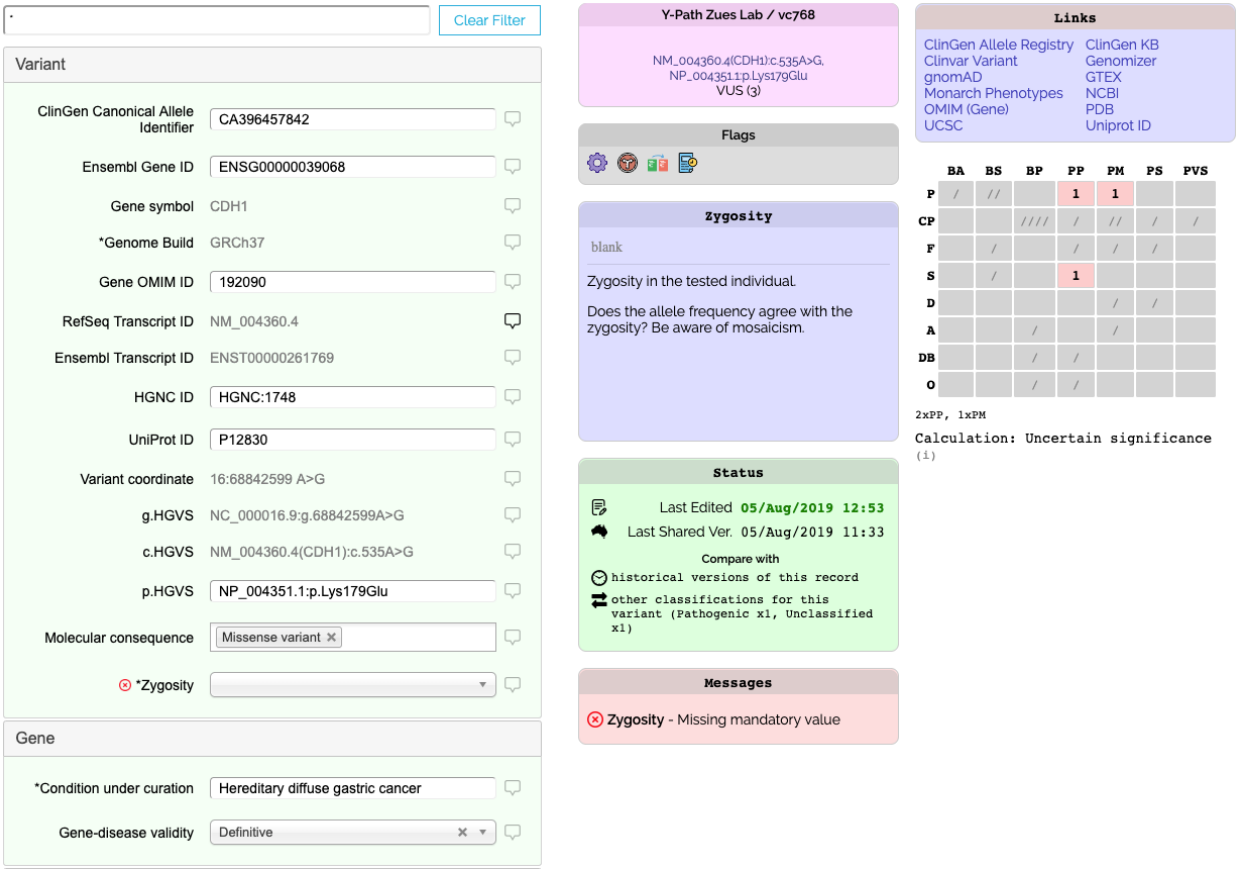
CHAPTER

THIRTYNINE

CLASSIFICATION FORM

The Classification Web Form can be used to create and edit classifications directly within VariantGrid.

39.1 View



To quickly see all fields that have values for a classification, enter "*" into the filter box at the top of the classification. To see all possible fields, enter "**" in the filter box. To find an individual field, start typing the label of the field into the filter e.g. "gnomad".

39.2 Identify Errors

A record might not be shared as there are outstanding validation errors. In the Messages box on the form it will list any errors. If possible fix those errors in your curation system and then they should be fixed on the next sync.

39.3 Change History / Diff

Each version of a record published in VariantGrid is recorded, by clicking on "Compare historical versions of this record".

If there are other classifications for the same variant, there will be a link to compare them there too.

39.4 ACMG Guidelines

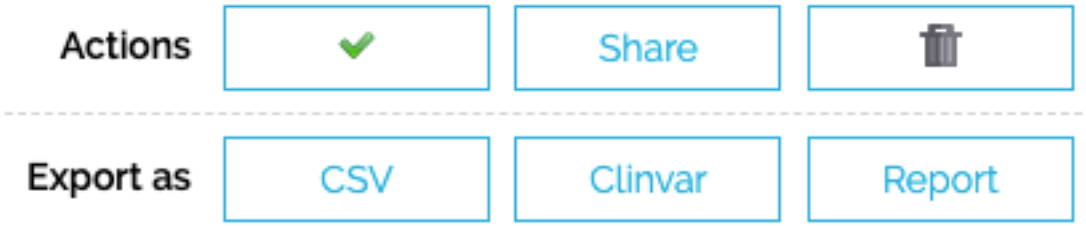
The classification form has fields for the ACMG Guidelines, e.g. PM4, BA1 - the meaning of each is given in the help. See Guidelines

VariantGrid displays a grid of ACMG fields with each row being a category of data, and each column representing the strength of evidence for benign or pathogenic.

- The number of met criteria for a given box will be shown as a number.
- Explicitly unmet criteria will show as "/"s.
- Criteria not yet marked as met or unmet will show as "?"s.

The various values will be plugged into the ACMG formulae and a recommended overall clinical significance will be displayed. This calculated value has no affect on any of the data, the user is still able to set the overall clinical significance to whatever (hopefully justifiable) value they like.

39.5 Actions



At the bottom of the form there will be a list of action buttons.


Tick - re-submits the classification at its current change level. For any manual changes to be seen, this button will need to be ticked.

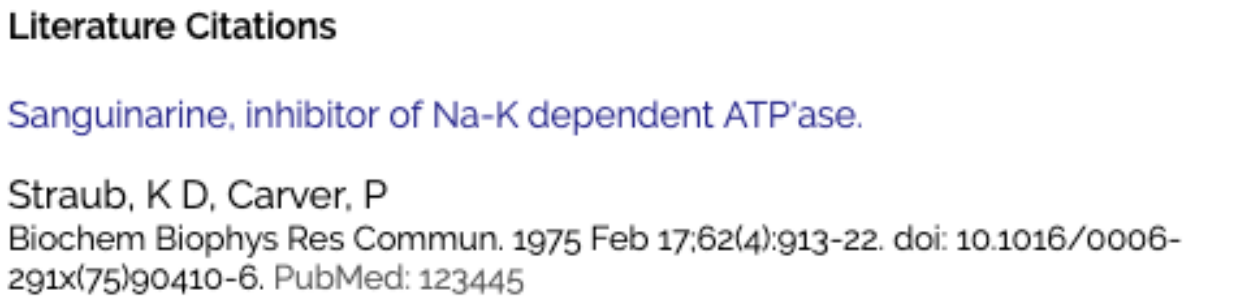
Share increases who can see the classification, see Classification Sharing


Delete/Widthdraw - Delete an unshared classification, or withdrawal (hide/ignore) a shared one.

39.6 Export

You can also export the single record as CSV, a preview of the Clinvar format or as a report. (The report does require that your lab has a report template pre-configured.)

39.7 Literature Citations



Any PMID references in the form of PMID:123456 from anywhere within the classification will be summed together and listed at the bottom of the classification.

CHAPTER

FORTY

CLASSIFICATION SHARING

40.1 How to share

A classification submitted without errors can have its visibility increased by clicking the 

Share Button at the bottom of a classification. You can increase share levels, which are:

- **Lab**
- **Organisation** (your research institute or company if they have multiple labs)
- **Shariant Users** (sends to Shariant)
- **3rd Party Databases** (sends to Shariant & to further public databases)

Share levels can only be increased, and each level also includes all previous levels, see Shariant doc on sharing. The last two levels mean classifications can be sent to external systems.

40.2 External systems

VariantGrid integrates with Shariant, the Australian Genomics Variant Classification Sharing Platform, which helps labs meet sharing best practices, and alerts them if another lab classifies a variant differently.

If enabled (currently clinical diagnostic only, not research servers), the system will regularly check for classifications with *Shariant Users* / *3rd Party Databases* share levels and automatically send them to Shariant.

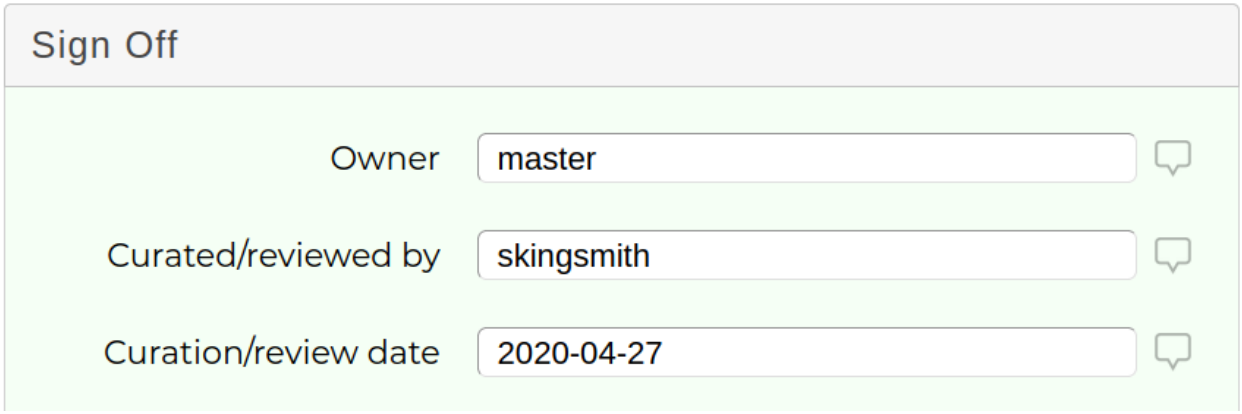
Warning: You can only increase a variant's share level, not reduce it (eg as someone may have seen and copied it)

40.3 Private fields

Some evidence keys have a "max share level" and are never shared beyond that level, regardless of the overall classification share level.

For instance **curated_by** and **curation_verified_by** have a max share level of institution, which means only your users can see them. Users from other organizations can see the classification was from your lab, but not who did the curation.

What your institute sees:



What others see:

40.4 Withdraw

You cannot delete a classification that has been shared, but you can "withdraw" it.

This will remove the record from most listings and search results, but will not remove it from any Discordance Reports that it had been involved in (it will no longer be a part of discordance calculations).

When a record has been withdrawn it can be unwithdrawn by clicking the same button (it should look like a rubbish bin with a raised lid now).

CHAPTER

FORTYONE

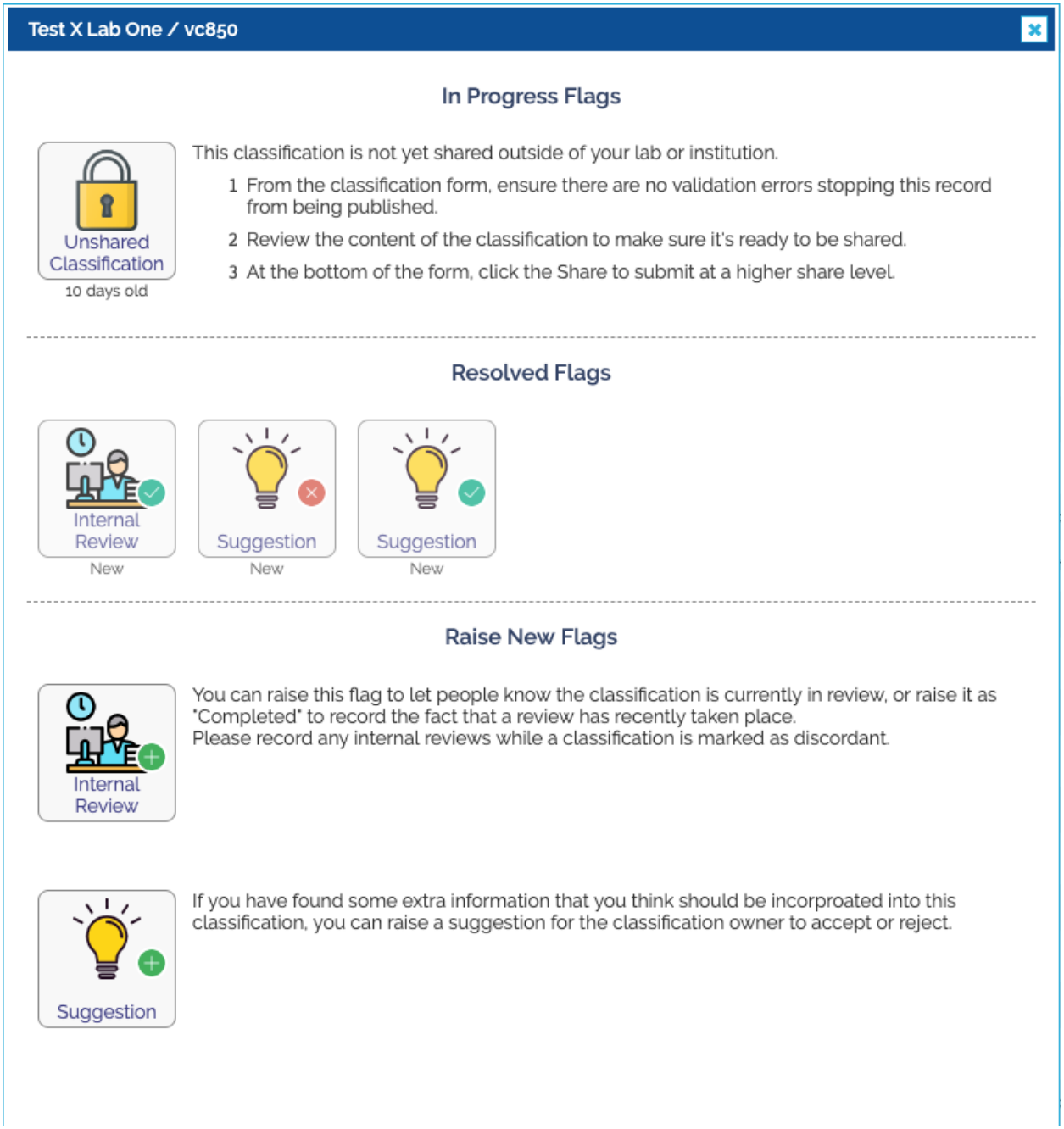
CLASSIFICATION FLAGS

Each classification flag indiciates that there is an action that needs to be performed against the classification.

Many of the flags will be automatically raised by Shariant, though some of them you will be able to open yourself.

To look at the details of a specific open flag, simply click on it to be taken to the flag dialog.

41.1 Flag Dialog



From the flag dialog you can view summaries about what flags are currently open, see a list of flags that have been resolved as well as raise new ones. Note that only important flags still show up when closed, e.g. suggestions and internal reviews and a few others.

In the provided screenshot we can see we have an open flag asking us to share the classification, a completed internal review, an accepted suggestion and a rejected suggestion, as well as the buttons to create new internal reviews and suggestions.

You can visit the details of an open flag, or a closed one by clicking on the icon.

From the details page of an open flag, depending on the type of flag, you can add a comment and potentially change the status of a flag.

You can raise a new flag by clicking on one of the icons near the bottom with a plus button.

(The kinds of actions you can take on flags will depend on if you're looking at a classification from your lab or another lab.)

See below for flags and how to solve them:

41.2 Flag Types

41.2.1  Discordance

This classification is in discordance with one or more classifications.

1. Ensure that you have completed an internal review of your lab's classification recently (within the last 12 months is recommended). If not, raise the internal review flag and complete an internal review of your lab's classification.
2. Review any outstanding suggestions against your lab's classification.
3. View the other classifications in the discordance report and view the evidence differing between multiple records via the diff page. If appropriate, raise suggestions against other lab classifications.
4. This Discordance flag will automatically be closed when concordance is reached.

This is discussed in the Classification Discordance page.

41.2.2  Internal Review

This classification is marked as currently being internally reviewed.

1. Once the internal review is complete, ensure you update the classification in your curation system.
2. Mark the internal review as Completed.

This is discussed in the Classification Discordance page.

41.2.3  Matching Variant

This variant has not been seen in this system previously. It should be linked to a variant given time.

41.2.4  Matching Variant Failed

We were unable to normalise the variant provided based on the c.hgvs and genome build values.

1. Please contact Shariant support for help in resolving this.

41.2.5  Outstand Edits

Edits have been made to this classification that are not included in a published version.

1. From the classification form, ensure there are no validation errors stopping this record from being published.
2. At the bottom of the form, click the tick to submit the outstanding changes.

41.2.6  Significance Changed

This classification has changed it's clinical significance compared to a previously published version.

1. Set the status of this flag to reflect the primary reason behind the change in classification.
2. Please also add a comment providing some context.

This is discussed in more detail on the Classification Discordance page.

41.2.7  Suggestion

Someone has raised suggestion(s) against this classification.

1. Review the contents of each suggestion.
2. If appropriate, make changes in your curation system and mark the suggestion as Complete.
3. If you decline the suggestion, mark it as Rejected.

41.2.8  Unshared Classification

This classification is not yet shared outside of your lab or institution.

1. From the classification form, ensure there are no validation errors stopping this record from being published.
2. Review the content of the classification to make sure it's ready to be shared.
3. At the bottom of the form, click the Share to submit at a higher share level.

41.2.9  Withdrawn

This classification has been marked as withdrawn. It will be hidden from almost all searches and exports.

1. If the classification is not of high enough quality or in error, you may leave it as “withdrawn” indefinately.
2. If you wish to un-withdraw the classification, click the open bin icon in actions from the variant classification form. (Note you can’t open a Withdrawn flag, but you can Withdraw/Unwithdraw from the classification form)

CHAPTER

FORTYTWO

CLASSIFICATION REPORT

42.1 Running the report

To generate the report from a classification, open the classification and scroll to the bottom. You will see a button called "Report". Click on it and you will then be able to copy & paste the report contents into a document.

42.2 Configuring the report

The report can only be configured by admin users - see admin docs

CHAPTER

FORTYTHREE

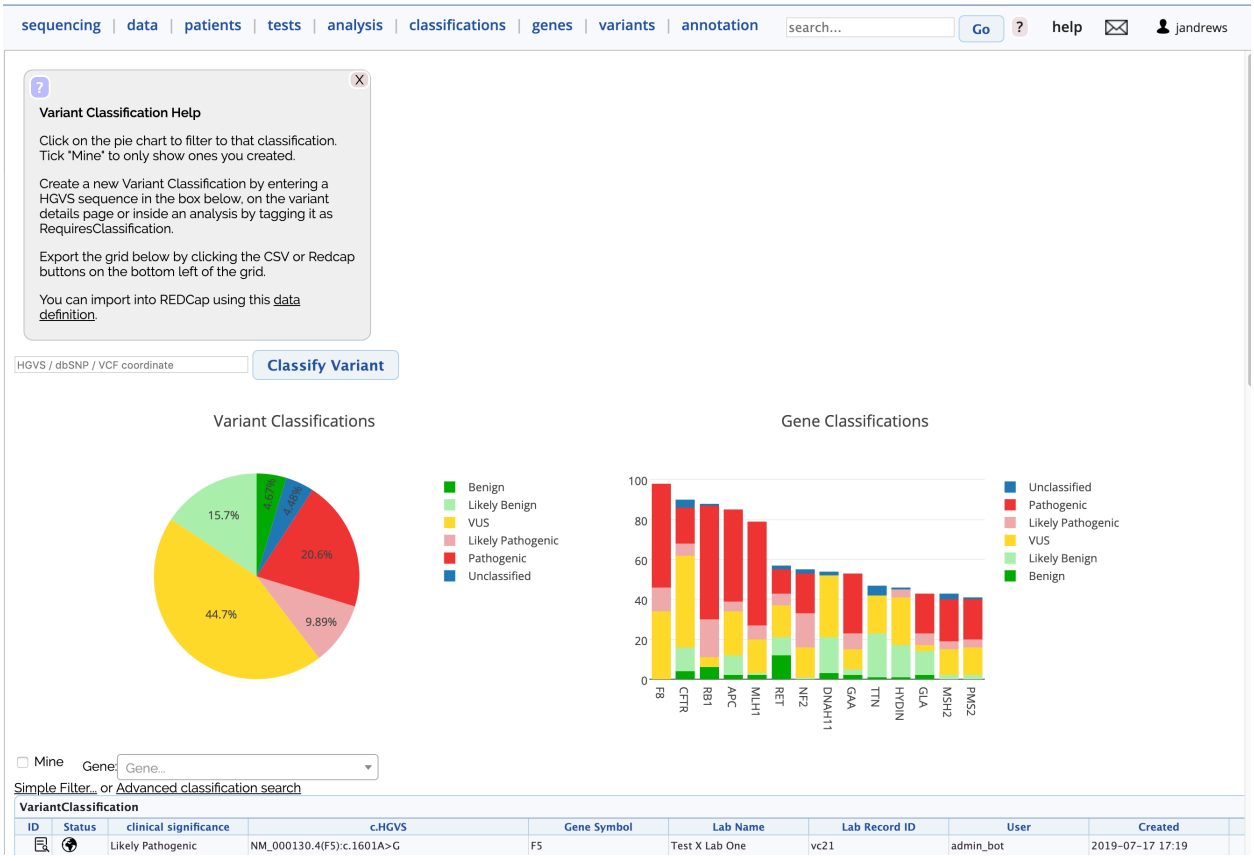
CLASSIFICATION REDCAP

Variantgrid supports the exporting of Variant Classification data into REDCap files. Note that this is currently the full extent of REDCap integration with Variantgrid, there is no support for importing REDCap records or exporting any other kinds of records in a REDCap format.

There are two parts to the REDCap export.

43.1 REDCap Definition

The data definition is available by opening the page help on the classification page.



The definition is dynamically generated from the variant classification evidence key configuration. We do our best to ensure that changes to evidence keys are backwards compatible for REDCap definitions.

The definition is laid out in such a way that up to 10 records can be grouped together in one record e.g. `vc_zygosity_1, vc_zygosity_2, vc_zygosity_3 up to vc_zygosity_10` This is so that variants for the same patient can be consolidated.

Note that the REDCap definition is primarily used as a read only representation of the data, doing large edits of data in REDCap is not recommended.

43.2 REDCap Rows

Important: Variant Classifications will ONLY be exported if `REDCap Record ID` has a value. All rows that do not have a value for `REDCap Record ID` will be ignored in the export.

At the bottom of the classification table there will be a CSV and REDCap download button. Clicking the REDCap download will download records that are:

- Available in the current filter (if the results are split over multiple pages all will be downloaded). For example if you filter to show "Mine" the records in the download have to belong to you.
- Have a value for `REDCap Record ID`.

Records that have the same `REDCap Record ID`, regardless of any other factors, will be grouped together as described earlier, re `vc_zygosity_1, vc_zygosity_2` etc

43.3 Technical Specifics

This means while single drop down fields work as you'd expect, multi-drop downs produce text that's harder to report on.

The evidence key definitions for selects have an explicit index for each drop down option. If adding more options (regardless of insertion order) a new index should be assigned and existing options should retain their index. This is to help keep newer REDCap definitions compatible with older REDCap records.

CHAPTER

FORTYFOUR

INDICES AND TABLES

- genindex
- modindex
- search