Shariant

Australian Genomics Health Alliance

Mar 11, 2024

WHAT IS SHARIANT

This documentation provides information both for web users or the product, and developers who are connecting to Shariant's API.

CHAPTER

ONE

SHARIANT FREQUENTLY ASKED QUESTIONS

1.1 What is Shariant?

Shariant is a controlled access platform designed to allow Australian laboratories and clinical services to: automate sharing of detailed and structured scientific evidence about clinically curated variants; communicate in real-time to resolve variant interpretation differences; and access gene- and disease-focussed expertise.

By sharing, key variant information will be provided to laboratories, and also clinical services, to help them better interpret genomic tests and improve the outcomes of testing for Australian patients.

To do this Shariant stores:

- Genomic coordinates of a curated variant
- Clinical significance of the variant (e.g. Pathogenic, Benign)
- Structured evidence that was used to determine the clinical significance
- Brief details of the condition and phenotype
- Technical details of the test used to detect the variant
- Contact details for the curating laboratory
- Updates that have occurred regarding a specific variant (e.g. from Shariant discordance resolution)

Shariant does/will not:

- Store genomic data files (e.g. BAM, VCF files)
- Replace a laboratory's existing curation tool

The information in Shariant is stored under controlled access, meaning it is limited to those who are experts in the field. Some of the information can subsequently be made public and shared with international databases, if approved by the submitting laboratory, to enable any patients undergoing genomic testing, worldwide, to benefit from the information.

During Phase 1, Shariant will provide many features to help laboratories manage these variants, including:

- Automated upload to Shariant and download to laboratory curation software
- Capture of detailed and structured evidence used to determine clinical significance
- An automated notification system to alert users to clinically important classification discordances
- An in-built communication platform to resolve classification discordances between laboratories
- Semi-automated submission of summary information to ClinVar to share internationally (upon laboratory approval)
- A web form for manual curation and input of a selected variant
- Checks to make sure that the variant naming and coordinates map correctly to international standards

- Audit trails to allow laboratories to monitor and trace user activity. It is anticipated that more features will be developed and made available in Phase 2 of Shariant. Please contact the Shariant team for more information (communications@shariant.org.au).

1.2 Why is Shariant being developed?

Decoding the clinical significance of DNA variants is complex and challenging, and is impeded by operation of clinical laboratories as data silos. It is currently difficult for one Australian laboratory to know if others hold evidence to support or refute the clinical impact of a variant they have assessed. Failure to share information leads to the risk of missing a diagnosis or misclassifying a variant.

Shariant is being developed to unite these data silos, and provide key variant information to laboratories, to help them better interpret genomic tests and improve the outcomes of testing for Australian patients. Additionally, Shariant will assist laboratories in meeting the National Pathology Accreditation Advisory Council (NPAAC) and Royal College of Pathologists of Australasia (RCPA) recommendations for submission of genotypic data to clinical databases.

Shariant aims to share information about clinically curated variants to:

- Combine siloed information and expertise for increased and faster diagnoses across Australian laboratories, for targeted treatment and better health care of patients and their families
- Reduce the risk of variant misclassification by allowing laboratories to identify potential interpretation differences before they occur
- Streamline communication between laboratories via an in-built communication platform
- Standardise and improve practices, information and terminology relating to genomic information across Australia, providing greater consistency in how we diagnose and treat diseases
- Facilitate knowledge transfer between the Australian clinical genomics community and international data sources, benefiting patients in Australia and globally.

Most importantly, Shariant is designed to assist healthcare professionals deliver better genomics results. If you have suggestions for how Shariant might be helpful to you, please let the Shariant team know (communications@shariant.org.au).

1.3 Who is behind Shariant?

Shariant is an Australian Genomics initiative, run under its national project on Clinical Variant Classification and Sharing led by Amanda Spurdle at QIMR Berghofer Medical Research Institute. The project was established in response to an identified need, from Australian Genomics' member organisations and Australian clinical genetic testing laboratories, for assistance in meeting the growing technical and administrative requirements for clinical data sharing.

The Shariant platform is based on existing software, 'VariantGrid.com', developed by David Lawrence at the Centre for Cancer Biology, an SA Pathology and University of South Australia Alliance.

The key personnel involved in Shariant are:

- Amanda Spurdle (Project Lead, QIMRB, QLD)
- Emma Tudini (Project Coordinator, QIMRB, QLD)
- David Lawrence (Lead Developer, CCB, SA)
- James Andrews (Software Developer, CCB, SA)
- Sarah King-Smith (Reclassification, CCB, SA)

Shariant is supported by both Australian Genomics and the Centre for Cancer Biology.

1.4 What is the advantage of submitting to Shariant over only submitting variants to ClinVar?

Shariant offers additional features to the international database, ClinVar, including:

- Collection, storage and sharing of detailed curation evidence against ACMG criteria in a structured format
- Sharing of this detailed curation evidence in a controlled access manner between Australian laboratories and clinicians first, bypassing potential issues of displaying data in a publicly accessible databases and sending data overseas
- A notification service alerting laboratories to potential discordances between laboratories or changes in classifications
- An in-built communication platform to assist in the resolution of discordances
- A central administrative node for submission to external databases (such as ClinVar) upon laboratory approval, reducing the technical and logistical burden associated with managing upload of data to databases and preventing laboratories from duplicating this process.

Shariant is also streamlining the process of submitting to ClinVar, enabling laboratories to do so in a semi-automated manner, at their discretion.

1.5 Who can access the data?

Shariant is a controlled access platform. Therefore, access to Shariant is restricted to experts in the field and managed via Australian Genomics. Users must agree to the terms of use, which includes clauses to forbid data re-distribution, and requires approval to publish findings using data obtained from Shariant.

Upon laboratory approval, clinically curated variants and associated evidence will be submitted to external databases (such as ClinVar). These external databases may be public (no login required) and can therefore be accessed by anyone with internet access.

1.6 Do we have to change how we perform variant curation?

No. Shariant simplifies the sharing of curated variants for laboratories, with little disruption to their current workflow. Laboratories can continue to use their existing curation systems such as Agilent Alissa, Golden Helix or a custom curation tool. Shariant will connect to each existing curation tool via an Application Programming Interface (API) to allow for automated submission to Shariant. Information shared in Shariant will then be available for download into a laboratory's existing curation system.

A web form is available to enter or edit classifications manually (for instance, if a laboratory wished to enter data by hand from paper records) but for many laboratories this is unnecessary.

1.7 How much does Shariant change my laboratory's clinical curation workflow?

Shariant maximises the benefits of sharing and discordance resolution, while minimising changes to workflows.

Sharing variant classifications and associated evidence is performed behind the scenes for most users, so that no additional steps are needed to share variants. The main difference is that classifications from other Australian laboratories will now be visible inside their curation system, so they will be aware when a variant has been seen before, and can re-use the expertise and effort of others. Laboratories will be sent discordance notifications if any variants they have classified have been given a different clinical significance by other laboratories.

1.8 How is data shared with my laboratory's curation system?

We work with your laboratory to build a small tool that runs on your intranet and automatically shares classifications between your curation tool and Shariant.

We are working with Agilent, Golden Helix and the pilot laboratories to build these connector tools. All programs built will be open source and available on GitHub, allowing laboratories to audit them. All communication will be driven by the client and the laboratories will deploy the client tools themselves, following internal security practices. For more information please see the Shariant help guide .

1.9 How does sharing data between different curation systems work?

Australian Genomics surveyed Australian NATA-accredited diagnostic genetic testing laboratories, and found almost all use the American College of Medical Genetics and Genomics (ACMG)/ Association for Molecular Pathology (AMP) or ACMG/AMP-like guidelines. Our data collection is based around these guidelines, and we collect:

- An internal laboratory identifier (to track or modify the classification)
- A way to uniquely identify a variant (HGVS nomenclature or genomic coordinates)
- Clinical significance (Pathogenic to Benign)
- Condition
- All 28 ACMG criteria (BS1, BS2, PM5, etc)
- Evidence used to support the evaluation. Sharing detailed curation evidence in a structured way gives Shariant an edge over other sharing tools. Comparing a variant's classification and associated evidence is as simple as looking at which ACMG criteria have been interpreted differently or whether extra hidden evidence exists.

1.10 What kind of evidence can I store against my classification?

Laboratories can submit any evidence they want.

The goal is to balance keeping evidence structured while also accepting all the different types of evidence that exist (e.g. databases, websites, bioinformatic prediction tools) - some of which haven't been released yet!

Thus, each piece of evidence requires a name, section, version and value (eg "gnomAD", "Population Data", "2.0.2" and "0.01"). For evidence names, laboratories should strive to re-use existing keys so that they match between laboratories. Sections are from the ACMG/AMP guidelines that group criteria into sections such as "Population Data", "Computational and Predictive Data" and "Functional Data". This helps to show the relevant data when comparing classifications between different laboratories.

For more information please see the Shariant help guide .

1.11 I work in a private laboratory, can my laboratory participate?

Yes. Shariant aims to connect to all interested Australian NATA-accredited genetic testing laboratories, regardless of whether a laboratory is public or private. Prioritisation of laboratories for connection to Shariant will be based on the following criteria:

- Interest of the laboratory
- Ease of connection to the laboratory's curation system
- Testing output of the laboratory
- Whether a laboratory is public or private

Please contact the Shariant team, if you are interested in connecting to Shariant (communications@shariant.org.au).

1.12 What is the timeframe for the development of Shariant?

Shariant is currently under rapid development, with new features being added each week. A beta-version of Shariant will be released for pilot testing in June 2019. After three laboratories have been successfully linked to Shariant, the platform will be rolled out nationally at all other interested Australian laboratories. Full technical support will be provided to assist laboratories connect to the platform.

Please contact the Shariant team if your laboratory has expertise in variant curation and is interested in testing the beta-release of Shariant.

1.13 How are Shariant and VariantGrid related?

VariantGrid is a variant database and web analysis platform, and is the technology licenced by Australian Genomics to build Shariant.

CHAPTER

TWO

TECHNICAL ATTRIBUTIONS

Shariant is built upon VariantGrid technology.

See VariantGrid Technical Attributions.

CHAPTER

THREE

TERMS AND CONDITIONS

3.1 End User T&Cs

- Shariant end user terms and conditions v3 (current)
- Shariant end user terms and conditions v2
- Shariant end user terms and conditions v1

3.2 Citing Shariant

If making use of data from Shariant Australia in a publication, please acknowledge Shariant using the following text -

> This study makes use of data generated by the Shariant Australia community. A full list of centres who contributed to the generation of the data is available via email at communications@shariant.org.au. Shariant Australia is supported by Australian Genomics Health Alliance (NHMRC grant 1113531).

CHAPTER

FOUR

IF YOU ARE VIEWING CLASSIFICATIONS, FIXING ISSUES, RESOLVING DISCORDANCES

4.1 Web Interface Overview

It is intended that the majority of classification data from your lab be automatically synced to Shariant. Shariant is not intended to be a place for large manual data entry. Conversely classifications in Shariant should be automatically downloaded to your curation system.

Given this, there are still plenty of uses for the web interface.

- Identify and fix issues with classifications from your lab.
- Resolve any classification discordances.
- Manually find classifications.

4.2 Classification Dashboard

The classification dashboard presents classifications from your lab that require some actions (aka classifications with open flags).

If you get regular emails about classifications requiring attention, it is in reference to the classifications that appear on the dashboard.

See Classification Flags for details.

4.3 Classification Diffs

Configure columns	Y-Path Zues Lab / vc768 24/Jun/2019 11:43 highlighting diffs	Y-Path Hemes Lab / vc769 20/Jun/2019 16:55 highlight diffs
BA1	Not Met	Not Met
BS1	Not Met	Not Met
BS2	Not Met	Not Met
PM2	Pathogenic Moderate	Pathogenic Moderate
PM2 note	Absent from 1000 genomes, ExAC, etc.	Absent from 1000 genomes, ExAC, etc.
PS4	Pathogenic Suppoprting	Pathogenic Suppoprting
PS4 note	Multiple family members with gastric cancer in ex cel sheet	Multiple family members within 1 gastric cancer fa milies from internal database
Summary	The c.535A>G (p.Lys179Glu) variant is absent in the gnomAD cohort (PM2; ht tp://gnomad.broadinstitute.org).This v ariant has also been reported in at lea st one family meeting HDGC clinical c riteria (PS4_Supporting; internal labor atory contributor).	The c.535A>G (p.Lys179Glu) variant is a bsent in the gnomAD cohort (PM2; htt p://gnomad.broadinstitute.org).This va riant has also been reported in 1 famil y meeting HDGC clinical criteria (PS4_ Moderate; internal laboratory contribu tor).
Computational and predictive data		
CADD	29.0	29.0
fathmm	Tolerated	Tolerated
Grantham Score	56	56
MutationTaster	Probably Damaging	Probably Damaging
Poly-Phen2	Probably Damaging	Probably Damaging
SIFT	Damaging	Damaging
GERP	5.78000020980835	5.78000020980835
PhyloP	2.33299994468689	2.33299994468689
PVS1	Not Met	Not Met
BP4	Not Met	Not Met
PP3	Not Met	Pathogenic Suppoprting
PP3 note	Splicing analysis estimates the disruption of an exo nic ESE site but ESE predictions not well validated	Splicing analysis estimates the disruption of an exo nic ESE site

The diff page allows you to compare several classifications, be it the history of a single classification or comparing classifications within a clincal group.

It is a handy tool for resolving discordances as you can see exactly what's different between the classifications.

Do note that different labs will be using different curation systems with different syncing abilities to Shariant - so there might be more data in their system than is shown in Shariant.

From the diff page, if you observe something worth discussing you can raise a suggestion flag against the classification in question.

4.4 Classification Discordance

4.4.1 Terminology

Allele

Shariant refers to "alleles" as a way to join logically equivalent variants in different builds together.

e.g. `GRCh37 1:40773150 G>A` is linked to the same allele as `GRCh38 1:40307478 G>A`.

Discordance is applied across genome builds on alleles, so if your lab is using GRCh37 and another lab is using GRCh38 and both labs have a classification that is logically referring to the same variant, then discordance calculations will consider both of those classifications together.

Clinical Groups

Clinical groups are used to subdivide classifications within an allele. In many cases there will be only be one "default" clinical group for an allele, but in the following scenarios you might need to make new ones:

- The conditions between the classifications is significantly different (e.g. "Tietz syndrome" vs "Carpenter syndrome")
- The transcripts are significantly different

In most cases the "default" clinical group per allele will be sufficient.

Discordance

Two (or more) classifications are considered discordant if they:

- Belong to the same allele
- Belong to the same clincal group within that allele (including the "default" clinical group)
- Are shared at the Shariant Users level or 3rd Party Database level
- Have clinical significances that fall into two or more of the following buckets (benign/likely benign), (VUS A/B/C), (likely pathogenic/ pathogenic), risk factor e.g. likely benign is discordant with VUS, but likely benign is concordant with benign.

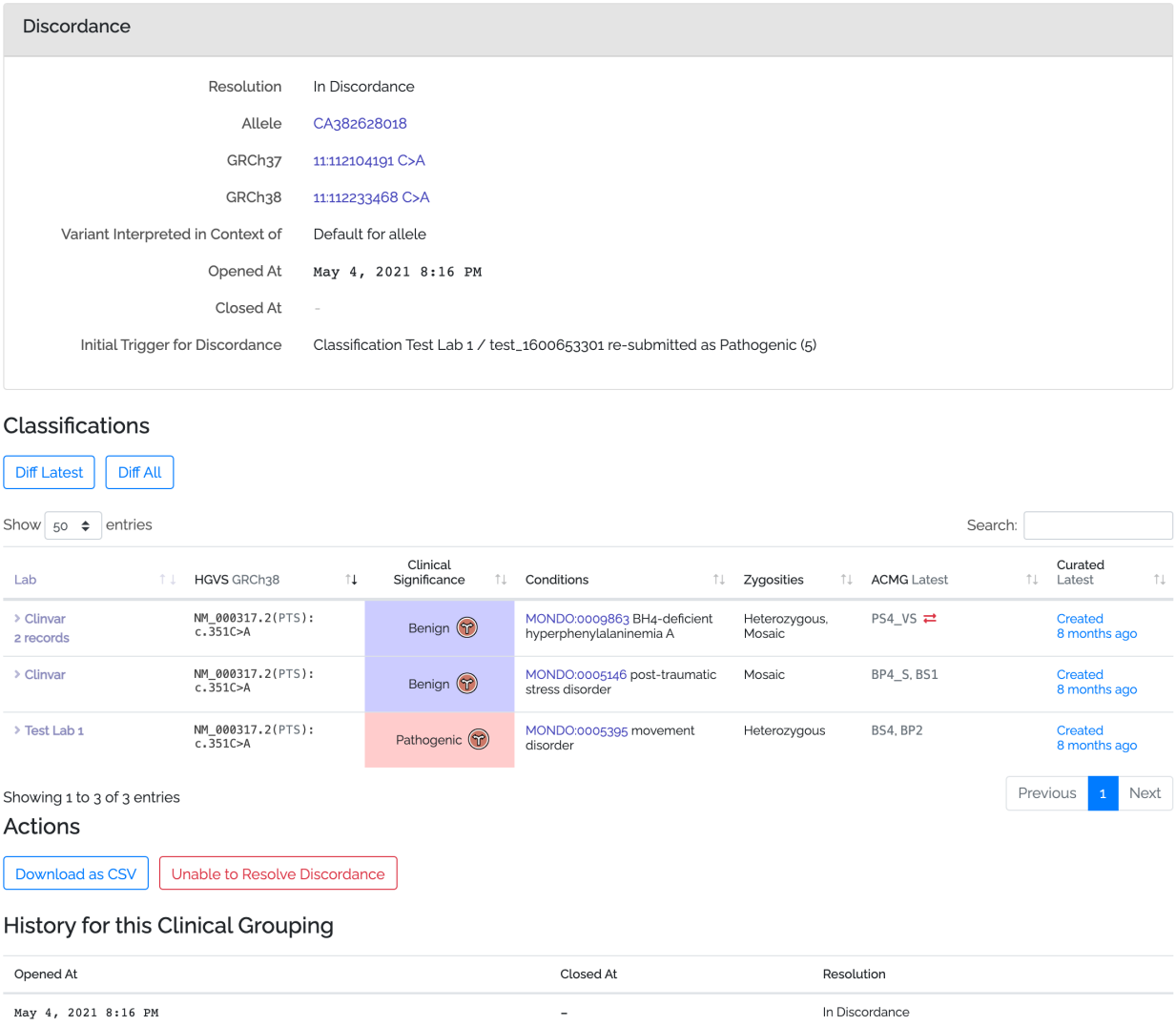
Transcript & condition do not directly affect the calculation unless a human determines the differences in those values designate a different clinical group.

4.4.2 Discordance... Now what?

So if one of your classifications meets the discordance criteria with another classification, what happens next?

4.4.3 Discordance Report

A good place to start is the Discordance Report.



This shows you:

- When the discordance was first detected and the action that triggerd it
- The records that were involved at the start of this discordance
- The clinical significance as it either is now (if discordance is onging) or how it was at the end of discordance.

Who's in Charge / Responsible

The owner of the classficaton that triggered the discordance is seen as the primary person in charge of resolving discordance. Note that this is not enforced in Shariant, anybody is free to take the lead resolving a discordance.

Internal Review

When discordance is first detected, it is recommended practice to perform an internal review of the classification (particularly if the classification hasn't been reviewed in the last 12 months).

You can raise an internal review flag in "In Progress" status and complete it later, or if you started and completed an internal review you can raise the flag as "Complete". This will then show up in the report that an internal review has taken place.

If during your internal review, data is found that should be changed, please change the record in your curation system and the update will be applied the next time your data in synced with Shariant.

Note that you are welcome to record internal reviews in Shariant that were performed outside of a discordance being detected, though doing so is not required.

Discordance Discussion

Once you've completed your internal review, you should investigate what evidence other labs have provided for the same variant.

The diff page is a good place to do this as it shows all the details of the classifications that are involved in the discordance discussion. (Diffs can be access from your classification, the discordance flag or the discordance report)

See details About the Diff Page here

After reviewing details from another lab's classification, you can raise  "Suggestion" flags on their classifications. "Suggestion" flags are able to provide a means of communicating with another lab.

When reviewing the other lab's classification, keep in mind what flags have already been opened and closed e.g. they may have not yet completed an Internal Review or there might be some previously recorded accepted and rejected suggestions. Also keep in mind that if a lab has agreed to make a change, it may take some time for that change to be synced back up to Shariant from the lab's curation system.

Change of Clinical Significance

As a result of an Internal Review, Suggestions or outside discussions you might decide to change your record's Clinical Significance (hopefully towards concordance). Update this in your curation system and later this will be synced with Shariant. A flag is then automatically raised asking for the reason behind a change. That reason will be recorded against the Discordance Report.

Concordance is Reached

If you or another lab:

- Change clinical significances so all the remaining classifications are in the same clinical significance bucket.
- Withdraw classifications that aren't in concordance. (see below)
- Sub-divide classifications into more meaningful groupings. (see below) Then you will have reached concordance and the discordance and discordance report will automatically close. Congratulations.

Withdrawing

If a classification that caused the discordance has been found to be an error and shouldn't be considered in its current form, the owner of the classification may mark it as Withdrawn from the classification form.

Withdrawn classifications are not considered during discordance calculations. (It will still appear on the report, but will be struck out to indicate it is no longer considered.)

Clinical Re-Groupings

If you determine you have two or more classifications that are discordant but for good reason, you can request an admin member change the Clinical Grouping of the records.

For example, if a variation is Pathogenic for "Tietz syndrome" and benign for "Carpenter syndrome", an admin can divide classifications up into both.

Continued Discordance

If after internal reviews and discussion, you have been unable to reach concordance, navigate to the Discordance Report page.

From here you can click "Unable to reach concordance". Please only do this as a last resort. Note that if there are any future changes to the Clinical Context (a new record is added or a record changes its significance) discordance will be re-opened.

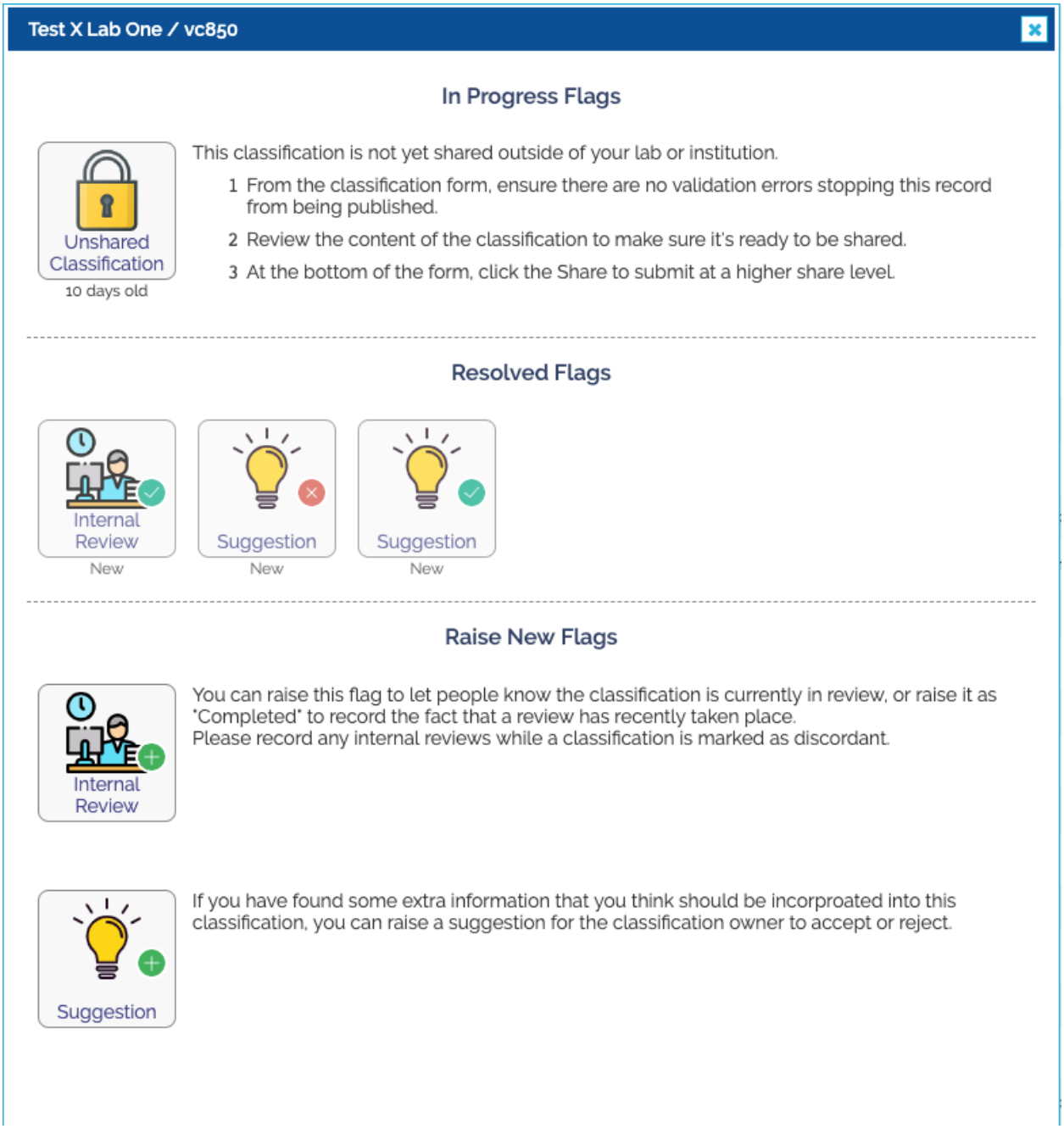
4.5 Classification Flags

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 you will be able to open some of them yourself.

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

4.5.1 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.

In the provided screenshot, we can see that 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 icons that have a "plus" button near the bottom of the dialog.

(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:

4.5.2 Flag Types

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.

Internal Review

This classification is marked as currently undergoing internal review.

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.

Matching Variant

This classification is not yet linked to a variant

1. If this has a status of In Progress we should match it to a variant shortly and no action is required.
2. If this has the status of Matching Failed, we were unable to normalise the variant provided based on the c.hgvs and genome build values. Please contact Shariant support for help in resolving this.

Matching Warning

The c.hgvs imported and the c.hgvs made after normalised by Shariant differs. This classification has been matched to a variant, but requires a manual check to ensure it was matched correctly

1. If you believe it was matched correctly, select a status of Variant Confirmed and Save
2. If you believe this is not the variant inteded, select a status of Variant Rejected and Save. An admin will then attempt to fix the problem.

Outstanding 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.

Significance Changed

This classification has changed its 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 add a comment providing some context.

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

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.

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" button to submit at a higher share level.

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" indefinitely.
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)

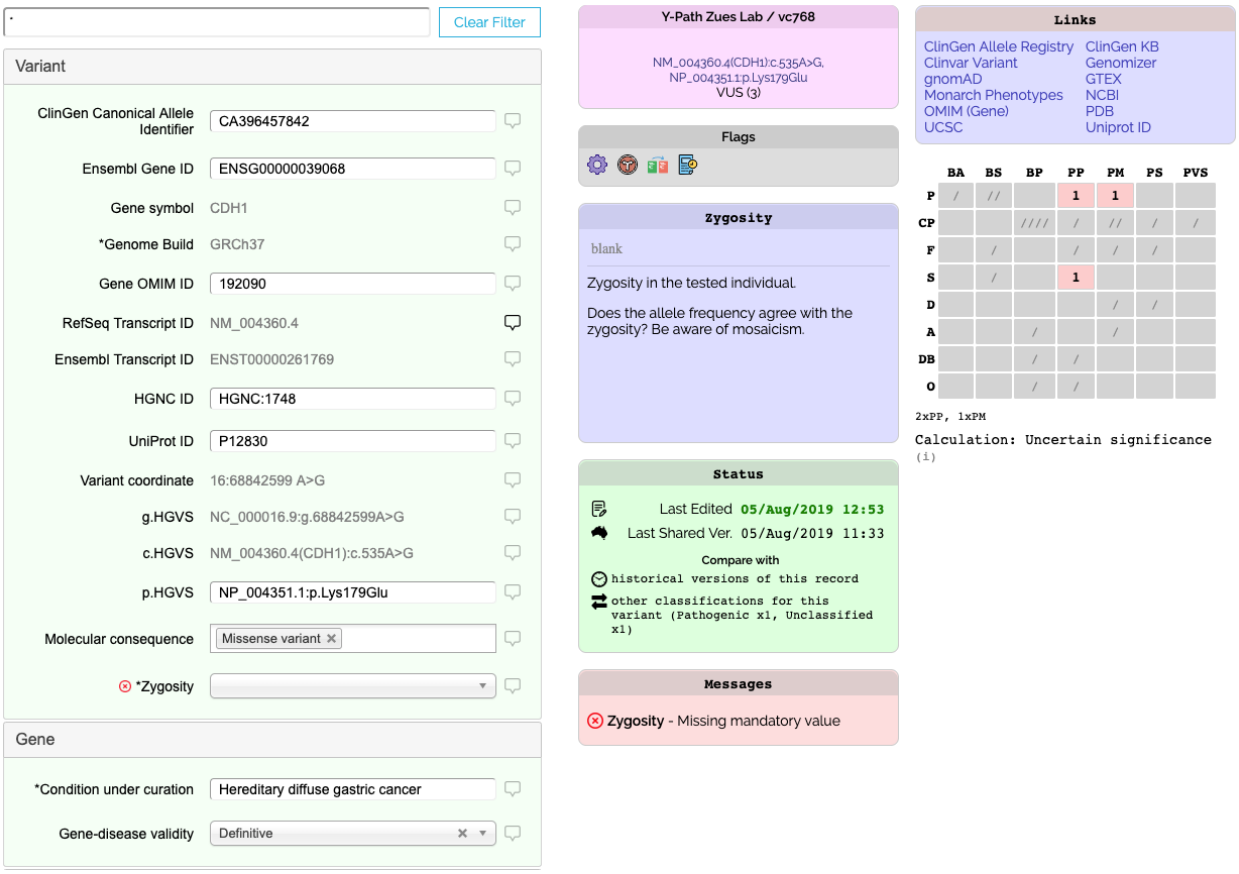
4.6 Classification Form

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

However it is expeced that classifications be created in your own classification system, and automatically synced to Shariant. That said, the Classification Web Form still provides some utility.

The web form provides a good readable version of Shariant's version of the web form.

4.6.1 View



To quickly see all fields that have values for a classification, enter "*" into the filter box at the top of the classification.

If you're looking at one of your own records, if a value was provided by your curation system to Shariant connector, it will be read only. If there are values that are never uploaded, you can edit them directly on the form. (This would

be a very time consuming process to do for all classifications so it's assumed it will only be done in special cases like discordances).

Note that for other users to see changes you have to perform a submit or share (see Actions).

You can also enter notes next to fields, be warned that if your connector to Shariant sends up notes, your notes might get overriden.

4.6.2 Identify Errors

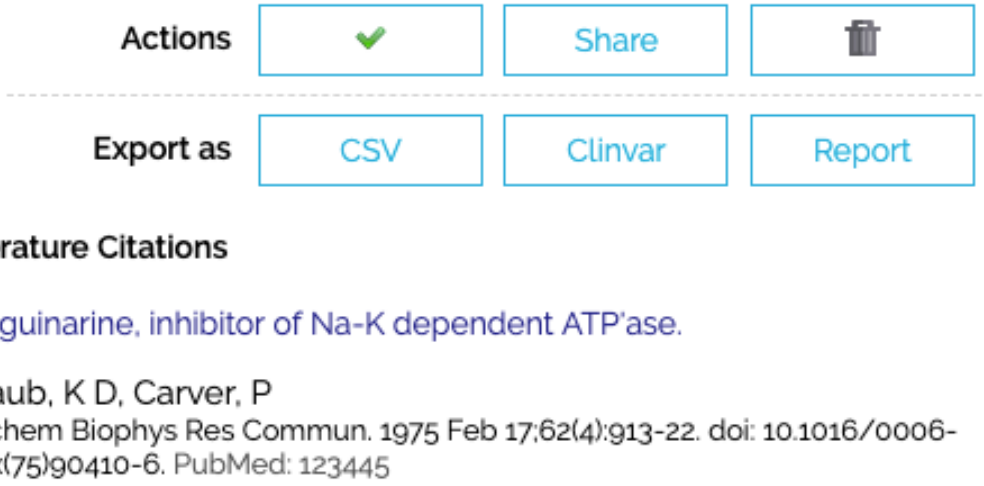
A record might not be shared as there are outstanding validation errors. In the Messages box on the form it will list any errors. If possible fix those errors in your curation system and then they should be fixed on the next sync.

4.6.3 Change History / Diff

Each version of a record published in Shariant 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.

4.6.4 Actions



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

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

Next to it is a Share button that allows you to increase the scope of who can see the classification. Important, increasing the Share level is not un-doable. The share levels are

- Just your lab
- Anyone within your organisation (if your organisation has multiple labs)
- All Shariant Users
- 3rd Party Databases (this will allow us to upload the record to Clinvar at a later date)

4.6.5 Delete / Withdraw

If the classification has only been shared at the lab or organisation level, you are able to perform a hard delete on the record. If it has been shared, instead you have the option to "withdraw". 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).

4.6.6 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.)

4.6.7 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.

4.7 Classification Listing

The classification listing page lets you filter through all the shared classifications in Shariant.

(!) You may see a field called "Create Classification", if you are having your classifications automatically synced avoid this.

Filter options

You can filter on Gene Symbol. Lab name (helpful to filter to your own lab to verify records) Flags - if you select multiple flags then records returned must have at least one of those flags. See Classification Flags for details.

Using Shariant's overall search (top right), you can still search for a specific variant.

4.7.1 Records Syncing Down

Hopefully as well as uploading to Shariant, your system is also configured to download records from Shariant and import them into your curation system. If this is the case it is recommended that you search for variants within your own classification system as you are probably well versed in its interface.

CHAPTER

FIVE

IF YOU ARE WORKING ON THE TECHNICAL CONNECTION TO SHARIANT

5.1 Integration Overview

5.1.1 Your variant classifications into Shariant

Goals:

- Ensure we can consistantly match records to variants.
- Compare classifications (across different labs) for the same variant in a meaningful way.
- Format the classification for submission to other databases such as Clinvar, which will have their own structure requirements.
- Avoid providing patient identifiable information.

5.1.2 Our variant classifications into your curation system

Goals:

- Ability to retrieve an extract from our Shariant and import it in a different track into yours and/or
- Be able to quickly use the web interface to search for relevant variants

5.2 Integration Getting Started

5.2.1 Step 1 : Defining the sync tool

Shariant accepts the upload of classifications via REST and is authenticated via OAUTH2. Shariant will not initiate any communication to your system (e.g. you wont have to open any ports or provide us authentication) instead it waits for uploads and requests for downloads.

You will need to have a tool running with access to your curation data on your network/cloud. This will be responsible for providing classifications to Shariant and if possible, automatically providing Shariant classifications as annotations to your system.

If your curation system is one that we've seen before, we can get you started with integration code. If it's new, we do have a Python library that will provide some structure around the calls. If your IT team already has a series of integrations, they're welcome to use any existing framework to perform the syncing.

5.2.2 Step 2 : Accessing and identifying data for the sync tool

Your curation system may have an API to access its data, it might have a database on your network that you can query or you might have to do a regular extract to a file and process that.

You probably don't want to upload all of your variants immediately, so it is useful to add some tags/fields to your system to help decide what to share and help with the integration.

- Last modified date
- Whether the variant is complete
- Share level

Last modified can be used to determine what you need to send to us since your last upload. The other fields are to determine whether to send the classification to us, and what the initial share status should be.

Some additional questions you may want to ask at this point:

- How do I uniquely identify a record from my lab (probably your internal database primary key)
- How frequently will the sync process take place?
- Will you re-upload all relevant records or only the ones that have changed?

5.2.3 Step 3 : Mapping from your structure to Shariants

Given any individual field you can enter in your system, how does it map to our pre-registered set of evidence keys? The sync tool may just need to use a different field name, or the mapping process may be more complicated.

For example, if your system has a Yes/No field called "This gene is known to be associated with X-linked recessive disease." that would map to our field "mode_of_inheritance" with a value of "x_linked_recessive".

Minimum mandatory set

Shariant enforces a small base level of required fields

- Fields to identify the variant, e.g. c.hgvs including gene symbol, ref and alt along with a transcript and a build (hg19 or 38)
- Lab record id: An ID you provide for the record so you can refer to it in future
- Clinical significance: How have you classified this on ACMG's scale of Benign to Pathogenic
- Condition: What condition are you curating against
- Zygosity: Zygosity in the tested individual.

Avoid personal identifiable information

Do not send us any information that could be used to identify the patient, specifically avoid names, detailed pedigrees, birth dates and addresses. You must also not enter such information into summaries or other fields you are mapping from your curation system.

Your **lab record id** should be a number like "10125" which doesn't mean anything and can only be matched to a sample or patient if you have access to your own secured patient record system or a record of the mappings.

5.2.4 Step 4 : Hosting the connector

Some labs will find it difficult to run arbitrary code on a machine in their domain without violating IT policies. Shariant maintains several s3 buckets where clients can upload files to be automatically mapped.

Discuss with the Shariant team if you think this will be a part of your solution.

5.2.5 Step 5 : Maintenance & user interaction

The Shariant API will return error messages for records that fail variant matching or are missing mandatory fields. Someone needs to periodically examine these logs and resolve any issues.

Individual classifications can be assigned to different users, if there's something on a classification that requires attention, the linked user will be notified by email. Some issues can be fixed by updating your curation system and waiting for the next sync, others will require interaction with the Shariant website.

Assigning resposible users may cause less coordination work for your team, though any lab member can see and edit all of your lab's classifications.

5.2.6 Step 6 : How best can your system integrate data from Shariant

Shariant provides an API for bulk downloading of classifications. Currently we provide classifications in our own JSON format, CSV, MVL, VCF. It is expected that most systems will be able to use these formats but some may require a custom solution.

5.3 Variant Classification Lifecycle

Here's a general view of what will happen to a classification. There are more details about each step.

5.3.1 Assign an internal id

Before a record is uploaded to Shariant, it should be given an alpha-numeric id that is unique to your lab. This "lab_record_id" can then be referenced in the future to refer to the classification.

5.3.2 Uploaded to Shariant

Along with your "lab_record_id" the sync tool will upload your evidence data to Shariant. If a record already exists with that id it will update the fields you provide, otherwise a new record will be created.

5.3.3 Variant matching

Using the fields you've provided, Shariant will start a process of variant matching for that classification. The record will be given a "Matching Variant" flag. Shortly after that flag will be automatically closed, if a variant couldn't be matched the record will now be flagged as "Matching Variant Failed" and the record will need to be deleted and attempted to be imported again.

5.3.4 Submitting/Sharing

As the sync tool uploads the records, it can also request that they be submitted to a specific share level. Share levels (covered elsewhere) determine who can see your record, including exporting to 3rd parties. Note that records must be free of errors to be submitted and shared.

5.3.5 Discordance

Once you've shared a record wider than your organisation, it will be included in discordance calculations. If another classification for the same variant (even from your lab) has come to a different clinical significance the records will be marked with a "Discordance" flag.

5.3.6 Updates

The same record can be uploaded an indefinate number of times. Shariant will store a version for each change, presenting the latest record to other users but previously submitted versions will still be accessible. As your curation system is considered the "source of truth", the cycle that includes re-uploading classifications is vital.

5.3.7 Clinvar Submission

TODO

5.4 Variant Classification Sharing

Shariant has the following concepts for handling who has access to what

A user can belong to multiple labs, though typically a user will only belong to one. Please contact us when a staff member leaves your lab and we can disable access to your lab's records.

Variant Classifications can be seen in two modes.

- The live editable copy
- A read-only version shared at a given point in time

If you or someone from your lab created a variant classification, you will be dealing with editable copy. If someone from outside your lab shares a record with Shariant, you will be dealing with the latest published version of that record. If they make changes and share it again, you will then have access to the new version. The same applies when you share records with Shariant.

Users with access to the editable version can elect to share the record in its current state as long as there are no outstanding validation errors. This will give other users read only access to the data as it is when the publish action was performed.

5.4.1 Share levels

Sharing can be done at several levels. Each level encompasses the level before it, and once it's shared at a certain level it can only be shared at that level or higher in future. The share levels are:

See the Sharing section in the API for information on how to utilise these share levels.

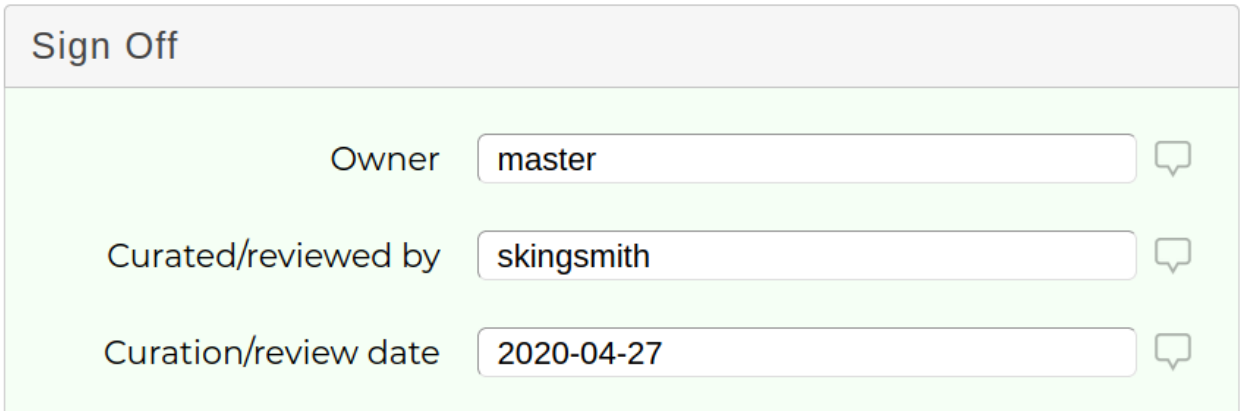
Just a reminder that the purpose of Shariant is to share records.

5.4.2 Evidence max share level

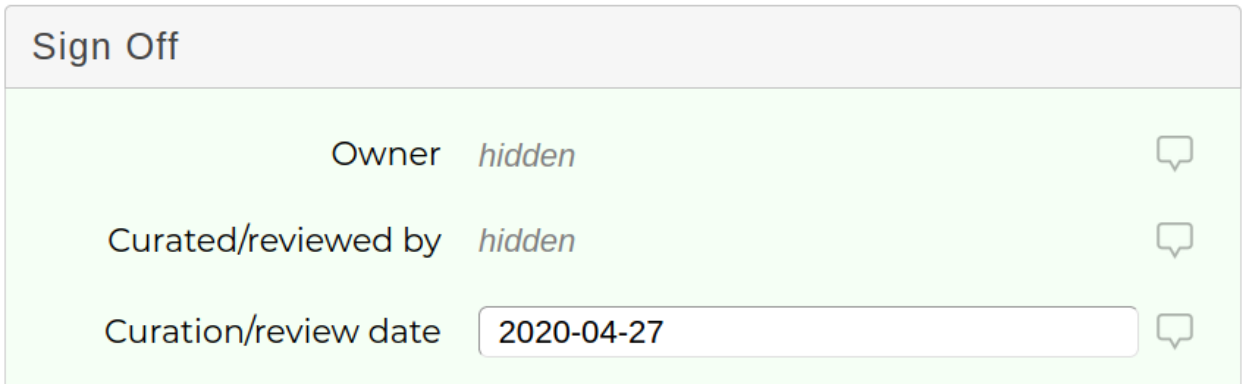
Some evidence keys have a "max share level" and are never shared beyond that level, regardless of the overall classification share level.

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:



5.5 Variant Matching Overview

5.5.1 Abbreviations

5.5.2 Introduction

A variant is composed of the reference, position within the reference, and the base change. Examples include HGVS [1] and the chrom/pos/ref/alt fields of a VCF file [2]. Below shows the same variant represented in different ways:

It is critical to assign variants to a unique identifier, so that it is possible to link variants from different sources, such as external population databases like gnomAD, to track observations from different VCFs, or classifications from different labs.

5.5.3 Difficulties and solutions

VCF or HGVS g./c. do not specify the transcript

Genomic coordinates (such as VCF or HGVS g./c. notation) can uniquely resolve base level changes, but may have different protein changes depending on the transcript used. The choice of transcript is significant for curation (e.g. splicing, pathogenicity prediction tools). There is the concept of "canonical transcripts" (a default transcript used when no transcript is specified) but canonical transcripts differ between labs and may change over time. Additionally, intergenic variants may not have a transcript.

Shariant Solution: It is strongly recommended that classifications submitted with VCF or HGVS g./c. Evidence Keys, also include the versioned transcript identifier used for curation. Classifications without a transcript will be linked, but a warning will be provided if a variant is inside a gene.

5.5.4 Ambiguous build or transcript versions

Not all VCF files specify the genome build. Some labs have stored classifications using VCF coordinates in custom databases but have not specified a genome build. Additionally, it is not always possible to check the genome build based on the variant submitted.

Labs that use HGVS coding DNA notation often store the transcript (e.g. NM_003607:c.325A>G) but not the transcript version (NM_003607.3:c.325A>G). While unlikely, it is possible that a variant in a different version of the same transcript may refer to different base changes.

Shariant Solution: Submission of a variant without a genome build being specified will not be allowed in Shariant and the record will trigger an error. Additionally, submission as HGVS g./c. without a transcript version will be strongly discouraged. If a laboratory does not have a transcript version, then the latest version of a transcript will be used to load the variant, and a warning will be provided that this may not be correct.

5.5.5 HGVS protein changes

Some medical scientists work with HGVS protein expressions. Even if the transcript is specified, codon degeneracy (redundancy of base changes leading to the same amino acid) means the nucleotide-level change cannot be distinguished. Determining the specific nucleotide change may be necessary (e.g. splicing).

Shariant Solution: HGVS protein changes are collected as Evidence Keys in Shariant, but are not used to link a classification to a variant. A laboratory will be unable to submit a variant with the protein change alone.

5.5.6 Multiple representations within the same build

It is possible to describe a variant involving an insertion and/or deletion in different ways for the same build or transcript versions. For instance, if the reference is "GGG" and a single G is deleted, this variant may be described as deleting either the first, second or third G.

Conventions have appeared, such as left aligning variants (Tan et al [3] - which introduces a tool vt normalise).

Regardless of the convention chosen, this must remain consistent within a laboratory's curation system and, in this case, within Shariant.

Below shows a table (from ClinGen Allele Registry [4]) of left and right aligned variants:

Popular variant callers (such as GATK) produce VCF files which are not normalised, though it is quite common for pipelines to use VT normalisation to left-align VCF files.

HGVS nomenclature specifies right alignment. The table below (copied from ClinGen Allele Registry [4]) shows multiple HGVS expressions for the same change.

Shariant Solution: HGVS coordinates are converted into VCF coordinates, then all VCF (or VCF from HGVS) coordinates are run through VT normalise before linking to a variant.

5.5.7 Conversion between genome builds

It is not always possible to lift-over all variants to another build, with rates of ~97% being commonly reported.

Newer builds have resolved some difficult sequences and introduced additional haplotypes (different reference sequences for regions of the genome that vary between human populations). Therefore, it may be possible that two distinct variants in one genome build will lift over to become a single variant in another genome build, or vice versa. A naive lift-over of two separate classifications from the same genome build may make those same classifications discordant on another genome build.

Shariant Solution: The ClinGen Allele Registry will be used to solve this problem by providing a globally unique ID (CAid) which can link variants across different genome builds. When a classification is imported which resolves to a novel variant in Shariant, an API request will be made to the ClinGen Allele Registry to retrieve or create an CAid for this variant.

The variant classification discordance process implemented in Shariant will work against these CAids (please refer to Shariant Technical Overview).

CAids can also be used as Evidence Keys to provide unambiguous linking of classifications to variants, and simplify submission to ClinVar. References

The HGVS for a classification may change after liftover, see Gene/Transcripts/Builds FAQ

5.5.8 References

[1] Dunnen, J. T., Dalgleish, R. , Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J. , Roux, A. , Smith, T. , Antonarakis, S. E. and Taschner, P. E. (2016), HGVS Recommendations for the Description of Sequence Variants: 2016 Update. Human Mutation, 37: 564-569. doi:10.1002/humu.22981

[2] The Variant Call Format (VCF) Version 4.2 Specification https://samtools.github.io/hts-specs/VCFv4.2.pdf

[3] Adrian Tan, Gonçalo R. Abecasis, Hyun Min Kang; Unified representation of genetic variants, Bioinformatics, Volume 31, Issue 13, 1 July 2015, Pages 2202–2204, https://doi.org/10.1093/bioinformatics/btv112

[4] Pawliczek P, Patel RY, Ashmore LR, et al. ClinGen Allele Registry links information about genetic variants. Human Mutation. 2018;39:1690–1701. https://doi.org/10.1002/humu.23637

5.5.9 Version History

5.6 Variant Matching Technical

TODO : This the exact field combinations Shariant requires with examples

5.7 Gene / Transcript / Build issues FAQ

5.7.1 Q. How do genes and symbols work?

RefSeq and Ensembl curate genes and transcripts, giving them stable identifiers (such as ENSG00000139618) and versions.

Gene symbols (eg BRCA2) are decided by committees - eg HUGO / HGNC for human. A gene version has an assigned symbol, but this changes over time

5.7.2 Q. Why are some gene/transcripts only available in certain genome builds?

Transcripts releases are made frequently (RefSeq is on version 99 and Ensembl version 100 as of May 2020) but new transcripts are only aligned to the latest build. Here are build release dates:

So all gene/transcript versions released since 2013/06/28 are not available for GRCh37

Obsolete transcript versions are not mapped to new builds, so all gene/transcript versions replaced between 2009-2013 are available in GRCh37 but not GRCh38.

5.7.3 Q. Why do gene names change between genome builds?

A gene version has a fixed gene symbol, independent of build, eg ENSG00000164199 version 11 has the symbol 'GPR98', while in version 18 it is "ADGRV1" However, as per above, the versions of genes and transcripts available for a build will differ, so ENSG00000164199 is 'GPR98' in GRCh37 but ADGRV1 in GRCh38

5.7.4 Q. Why does the same gene/transcript version have different exons in different genome builds?

Transcripts are worked out as mRNA then aligned back against the genome builds to find the exon coordinates. If genome builds differ at that gene, for instance inserting or deleting sequence, or base changes altering splice sites, then exon lengths or ends can change.

5.8 Evidence Keys Overview

A variant classification in Shariant is mostly comprised of evidence key values.

Evidence keys can be viewed here

Extra details

- Evidence keys are flat, e.g. a key will not be nested within another key.
- Each key is of a certain type, see Evidence Key Types.

- Each evidence key allows for a free text "note". Use notes to include extra details, especially when the evidence key itself has to conform to a certain format, e.g. `{"bp2":{"value": true, "note": "Was 50/50 on this one"}`
- Each evidence key allows for a free text "explain". Use explains to include details about how your lab uses that field if it differs (or is more explicit) than the ACMG guidelines, e.g. `{"pm2":{"value": false, "note": "absent from gnomAD, other pop databases are not used"}`

5.9 Evidence Key Types

A key's value type will determine how it will behave:

5.9.1 (F) free-text

Free text, no default validation applies, avoid new line characters if possible.

Any HTML tags will be stripped out.

5.9.2 (T) free-text multi-line

Free text, no validation applies, can contain new line characters and the following basic HTML elements:

```
<br>
<b>
<i>
<ol>
<ul>
<li>
<u>
```

5.9.3 (S) select

The key will also contain an array of options. To submit a value, provide a value that matches an option's key. If the entry also has `allows_custom_values` then any value will be accepted.

5.9.4 (M) multi-select

As per the normal select, the key should also contain an array of options. To submit a value, provide either a json array or a comma separated string. The elements in the array or string should match the keys of options. If the entry also has `allows_custom_values` then any value will be accepted, and standard values and custom values can be intermixed. The ordering of values from submission will not be maintained.

When viewing records, the value will be an array of option keys, unless there were no values selected in which case the value will be `null`.

5.9.5 (B) boolean

Provide a json boolean value, or alternatively "true" or "false" case insensitive. Be aware that the web form is unable to distinguish between false and no value.

5.9.6 (D) date

Provide a string in the format of yyyy-MM-dd e.g. Jan 2nd 1997 would be "1997-01-02". Time parts are not supported.

5.9.7 (C) criteria

An ACMG criteria e.g. PVS1, PS1, PS2. Accepts values of

- `NM` for Not Met
- `BS` for Benign Strong
- `BP` for Benign Supporting
- `BA` for Benign Standalone
- `PP` for Pathogenic Supporting
- `PM` for Pathogenic Moderate
- `PS` for Pathogenic Strong
- `PVS` for Pathogenic Very Strong
- `true` for the default strength of of the criteria

5.9.8 (L) float

sorry (F) was taken accepts any number, if valid will be stored as a float.

5.9.9 (I) integer

Accepts any whole number, if valid will be stored as an int.

5.9.10 (N) unit

Accepts a number between 0 and 1, if valid will be stored as a float.

5.9.11 (P) percent

Has been deprecated in favour of always using unit. ~~Accepts a number between 0 and 100, if valid will be stored as a float~~

5.9.12 (U) user

Provide the email address of a user. Typically this will be for the owner of a record, in which case it's more important for it to be the user who will login to the system to fix the issues rather than the person who created the classification.

5.10 API Philosophy

There are some fundamental principals that we're currently applying to the API:

The connector sync code that uploads the data should be repeatable and predictable. When data is missing or something else needs to be fixed, rather than manually tweaking things via the API - make the adjustments to the data in your curation system then re-run the connector sync code.

In addition the above, your curation system is to be considered the source of truth for the data. Evidence keys referenced by the sync code will become read-only to web users - so web users can't change data only then to have it overriden by the next sync. Data changes should be done directly in the curation system ready for the next sync.

As long as the API submission is a well formatted JSON, nearly everything uploaded will create a new (or alter an existing) variation classification record. The record itself may be marked with errors but they can be fixed in a subsequent sync or interaction with the web UI.

There will be a level of normalisation applied to submitted data, "true", "TruE", "T", 1 will be converted to `true` - and "false", "F" and 0 to `false` for boolean fields, drop down field values will have their case corrected.

We'll be mainly accepting free form text, so citing of publications or other resources will be achieved by parsing through the text looking for patterns, such as "PMID: XXXX" rather than requiring special structure in the submission.

Under no circumstances should identifiable patient data be uploaded.

Please note there is a test server https://test.shariant.org.au/

Due to the terms and conditions, do not upload real data to the test server, instead use mock data in the same format as the production data.

5.11 API Authentication

The Shariant API will require authentication to perform any operation. The authentication is currently provided through OAuth2 via Keycloak.

The OAuth URL (for production and test) is `https://shariant.org.au/auth/realms/agha/protocol/openid-connect/token` provide a client_id of `shariant-client-tool`

Here's some sample Python code that will be able to login

```
from requests_oauthlib.oauth2_auth import OAuth2
from requests_oauthlib.oauth2_session import OAuth2Session
from oauthlib.oauth2 import LegacyApplicationClient

def ping() {
        client_id = 'shariant-client-tools'
        username = 'xxxx' #not an actual login, please subsitute with a username and␣
↪password we've provided you
        password = 'yyyy'
```

(continues on next page)

(continued from previous page)

```
        oauth = OAuth2Session(client=LegacyApplicationClient(
                client_id=client_id
        ))
        token = oauth.fetch_token(
                token_url='https://shariant.org.au/auth/realms/agha/protocol/openid-
↪connect/token',
                username=username,
                password=password,
                client_id=client_id
        )
        auth = OAuth2(client_id = client_id, token = token)

        r = requests.get('https://shariant.org.au/classification/api/evidence_keys',␣
↪auth=auth)
}
```

Note that any classifications you make through the API will be assigned to the user used to authenticate against the API by default.

5.12 API Variant Classification POST

All operations against a variant classification submission from your lab can be performed against a single end point: https://shariant.org.au/classification/api/classifications/v2/record/

Ensure that any posts there are passing the required authentication data and is sending data as ContentType "application/json"

5.12.1 Example Post

```
{
        "id":"xyz_pathology/north_street_lab/F03432",
        "data": {
                "c_hgvs": "NM_000071.2(CBS):c.1539C>T",
                "genome_build": "GRCh37",
                "clinical_significance": "VUS",
                "zygosity": "hemizygous",
                "bp4": true,
                "mode_of_inheritance": ["autosomal_dominant", "autosomal_recessive"],
                "literature": "Found a book PMID: 342244"
        },
        "publish": "institution",
        "return_data": "changes"
}
```

or

```
{
        "id":"xyz_pathology/north_street_lab/F03432",
        "delete": true
}
```

5.12.2 All Parameters

5.12.3 ID Part

To uniquely identify the record, its ID can be part of the URL, or provided as part of JSON submission.

While we will assign a Shariant ID to all records, upon creation of a record you can assign it your own internal ID that only needs to be unique within your lab. You can then continue to use your internal ID for all future references.

The ID part of a submission consists of the following parts:

e.g.

```
{"id":"xyz_pathology/north_street_lab/F03432.1557189685.485863"}
```

or

```
{"id":{
        "lab_id":"xyz_pathology/north_street_lab",
        "lab_record_id":"F03432",
        "version":1557189685.485863
}}
```

5.12.4 Versions

Any change to the evidence of a record will automatically create a new version of that record.

A version is denoted by the UNIX timestamp (with decimal places) of when the record was altered. If a value of version is provided in the ID part, then you will be referring to the read only version of the variant classification at that point in time and won't be able to perform any operations against it. Versions cannot be un-published after being published, but more up to date versions can be published to become the new default.

5.12.5 Evidence Operation

Evidence part can be provided under one of the following. A submission can only provide a maximum of one of these elements at a time.

create

Provide evidence for a new record. Will error if the ID part matches an existing record.

patch

Evidence included in this will be merged with the existing evidence of a record. Will error if the ID part doesn't match an existing record.

upsert or data

Will either act as create or patch depending on if the ID part matches an existing record.

It is suggested that you only submit with upsert, as it stops labs from having to keep track of if a record has already been submitted.

overwrite

Like upsert, except all existing evidence (if any) will be overwritten.

5.12.6 Evidence Format

The content of create/patch/overwrite/upsert will be using keys as seen in the evidence keys section, with keys matching a value or a dictionary with a key of value or note. Notes allow you to associate arbitrary text with an evidence key, especially useful for boolean or (multi)select fields.

The evidence itself should in the form of `key : value` or `key: { "value": value, "note": note}` e.g.

```
{
    "id": "...",
        "upsert": {
                "literature": "Found a book PMID: 342244",
                "affected_status": {
                    "value": "yes",
                    "note": "Need to double check this"
                },
                "condition": {
                    "value": null
                },
                "search_terms": null
        }
}
```

How the data is provided is important for mixing the use of API and the web form. By default, any key with anything on the right hand side other than a straight `null` will be immutable on the web form. This is because your own curation system is deemed to be the source of truth about a variant classification. We want to avoid a scenario where users fix data on Shariant and then your curation system is out of date. This could be followed by another sync operation where the curation system's out of date data then upsets over the top of the correct data in Shariant. The selective immutability will allow web form users to provide values for keys that your curation system can't provide, if necessary. Importantly, a web user's ability to change the text for a "note" is not affected by the immutability status.

Formats

`key : null` This completely blanks out any value associated with the key. Value, note, explain will be blanked. Web form immutability will be set unless otherwise configured.

`key : <value>` This will set the value for a key as well as wiping any note. Web form immutability will be set unless otherwise configured.

`key : { "value": "<value>", "note": "<note>", "explain": "<explain>" }` If only a subset of value, note or explain are provided, this will merge with existing data. e.g. only providing value will leave any existing note and explain untouched. Immutability will be set.

The preferred method is `key: {"value": x}` (with note only if your curation system can send notes). For records that don't have values for certain keys that you would normally sync, provide `key: null` instead of omitting the entry all together. This is so immutability is set appropriately.

Values

The individual entries valid for values will depend on the associated key types, e.g. "BP4" accepts "NM", "BS", "BP", , true

"variant_type" accepts "indel", "splice_site" etc

"mode_of_inheritance" accepts ["autosomal_dominant","other"] or "autosomal_dominant, other"

See the Evidence Key Types page for more details.

5.12.7 Sharing

You can provide a publish flag in a POST. If create/patch/overwrite/upsert is provided, the publish will relate to the record as it is after applying that change. Note that you cannot un-publish a version, just publish more up to date versions.

To publish include a value for publish in your JSON body. e.g.

```
{
        "id": "67",
        "publish": "institution"
}
```

institution

Visible to any user that belongs to a lab that belongs to the same institution as the lab the record was created against.

logged_in_users

Visible to all logged in Shariant users and will be included in Shariant exports to other labs around Australia.

global

Allowed to be shared with Clinvar or other 3rd party systems.

Each share level is inclusive of all previous share levels. If on Monday you published a record to level 3, then on Tuesday you published the same record to level 2 - general Shariant users will have read-only access to the record as it was on Monday, but users within your institution will have access to the more up to date version as it was on Tuesday.

5.12.8 Deleting / Withdrawing

Records that haven't been shared with logged in users or global can be deleted by including `"delete": true`

Records that have been shared with logged in users or global can be withdrawn by including the same `"delete": true`

Withdrawn records: Can be "unwithdrawn" with a subsequent submission for the ID that doesn't have `"delete": true` or via the dashboard within the application. Do not show up in exports. Do not show up in classification searches (unless searching for the ID). Do not count towards discordance calculations. Will still be accessible using their ID, and will still appear in historical discordance reports that they were previously involved in.

5.12.9 Bulk POST

To perform multiple operations in one call, simply post a JSON array instead of a dictionary, e.g.

```
{
        "records":[
                {"id":77, "patch": {"ps1":"NM"}},
                {"id":78, "delete": true},
        ]
}
```

Though it is best if this is batched to no more than 20 operations per POST.

5.13 API Variant Classification Owner

A classification is always assigned to a single user, as well as a lab.

In addition to the internal user of a classification, there's an evidence key called "owner". If the value provided for "owner" is valid per the criteria below, the user of the record will be changed to it. If no value is provided for "owner", the "owner" key and the internal user will both be set to the user who created the new classification (be it done over the web interface or using the API).

Criteria for owner:

- The value must be an email address.
- The email address must match an existing Shariant user.
- The user must have logged into Shariant at least once.
- The user must belong to the lab that the record is assigned to.

If a value for owner is provided that doesn't match the above criteria, a warning will be raised but there will be no other adverse effects.

5.14 API Variant Classification JSON

Retrieving classification data by JSON is possible, but it is expect that the Variant Classification Download will offer a format more friendly to your curation system e.g. VCF, MVL.

If the format you need is not available, contact us and we'll see if we can add it.

An example of the result of a GET or POST below


```
{
    "id": 31,
    "meta": {
        "lab_record_id": "x77",
        "institution_name": "Example Pathology",
        "lab_id": "example/unit_1",
        "lab_name": "Example Unit 1",
        "title": "Example Unit 1 31",
        "version": 1544864360.852882,
        "can_write": true,
        "can_write_latest": true,
        "flag_collection": 13,
        "has_changes": true,
        "last_edited": 1553238219.242146
    },
    "allele": {
        "clingen_allele_id": "CA396457842",
        "genome_builds": {
            "GRCh38": {
                "variant_coordinate": "10:87957955 C>T",
                "g_hgvs": "NC_000010.11:g.87957955C>T",
                "c_hgvs": "NM_000314.4(PTEN):c.737C>T",
                "variant_id": 3326581
            },
            "GRCh37": {
                "variant_coordinate": "10:89717712 C>T",
                "g_hgvs": "NC_000010.10:g.89717712C>T",
                "c_hgvs": "NM_000314.4(PTEN):c.737C>T",
                "variant_id": 2414891
            }
        }
    },
    "publish": "logged_in_users",
    "data": {
        "condition": {
            "value": "OMIM:219700",
            "db_refs": [
                {"db": "OMIM", "id": "OMIM: 219700",
                "idx": "219700", "url": "http://www.omim.org/entry/219700",
                "summary": "CYSTIC FIBROSIS; CF"}
            ]
        },
        "gnomad_af": {
            "value": 0.002,
            "explain": "gnomAD 2.1.1"
```


(continues on next page)

(continued from previous page)


```
        },
        "c_hgvs": {
            "value": "NM_000314.4(PTEN):c.737C>T"
        },
        "g_hgvs": {
            "value": "NC_000010.11:g.87957955C>T"
        },
        "sample": {
            "value": "blood"
        },
        "zygosity": {
            "value": "heteroplasmic",
            "note": "xyz"
        }
    },
    "messages": [
        {
            "key": "clinical_significance",
            "code": "mandatory",
            "message": "Missing mandatory value",
            "severity": "error"
        }
    ],
}
```


5.14.1 Top Level

5.14.2 Allele

5.14.3 Data

5.14.4 Meta

5.15 API External References

Shariant parses all submitted values and note fields looking for external reference patterns (such as PMID). e.g.

"See the paper PMID:123456 about this"

Would parse out a reference to PubMed


```
{
"value": "See the paper PMID:123456 about this",
"db_refs": [{
    "db": "PubMed",
    "id": "PubMed: 123456",
    "idx": "123456",
    "url": "https://www.ncbi.nlm.nih.gov/pubmed/?term=123456",
    "internal_id": 50165
}]
}
```

These values will be formatted into a specific structure on download, but for upload they just need to be in the format of REF_TYPE:REF_NUMBER.

There is some leeway on the formatting, e.g. for PMID the following are allowed:

- PMID:123456
- PMID123456
- PMID 123456
- PMID#123456
- PMID 123456,6543321
- pmid 123456

Typically it's the case insensitive key term followed by any number of spaces, hashes or colons and then the referenced ID.

Some terms do allow a comma separated list of values, but this is not preferred.

As of October 2019 the following are supported:

5.16 Variant Classification Downloads

5.16.1 Overview

As well as having your lab upload classifications to Shariant, we want to get Shariant's classifications in your curation system so that you can quickly see if a variant has been previously classified using your regular workflow.

The download file can be downloaded by a user manually, or streamed directly into your connector with additional coding, and can be in one of the following formats:

- CSV
- JSON
- MVL
- VCF

More detail about each format is provided below.

Within each format there are some configuration options. Please discuss with the Shariant team if none of the formats suit your needs as we can develop additional formats or tweak existing ones.

5.16.2 Configuring Your Download

If you navigate to https://shariant.org.au/variantclassification/export you are presented with the download builder.

This page lets you choose if your downloads are going to be done manually in a web browser or invoked via API, the genome build, which labs to include/exclude, the format and any additional format parameters. (Via web browser or API result in the same file, but the web browser link will provide you with a login page if you're not currently logged in)

Once configured, a re-usable URL will be produced that you can either bookmark or invoke from custom connector code. Some formats provide all the data Shariant has, while others only have summary data and a link to the relevant Shariant data.

5.16.3 CSV

- The CSV format provides one row per variant classification and all evidence other labs have uploaded.
- The first few columns are header information, e.g. record ids and the normalised c.hgvs value. These columns are then followed by columns of imported data.
- The CSV can be configured to use codes, or more human readable headers and values.

5.16.4 JSON

The JSON format is what Shariant internally uses when displaying data on web pages and closely matches the internal database representation of records.

- JSON downloads will download an array called "records" with one JSON object per variant classification.
- Each record classification will show evidence in the format of "key": {"value": X, "note": Y} etc.
- Each record will have an allele section that will show data about the variant lifted over to GRCh37 and GRCh38.

5.16.5 MVL

The MVL format is used for Agilent's Alissa Interpret interpretation system.

- For Alissa 5.2, the MVL will need to be sent to Agilent for processing. If using Alissa v5.3, this can be uploaded manually in the Alissa interface.
- The MVL format provides one entry per variant/transcript combination, but that entry will include summary details and links about all classifications.
- The default settings for the MVL will work with most instances of Alissa, but it's important to check which clinical significance values are accepted by your instance (e.g. LIKELY_PATHOGENIC vs Likely Pathogenic).
- One important detail is that if there are classifications for the same variant that result in different clinical contexts, we can only provide a single overall value - so you can choose if that value is the least or most pathogenic. It's recommended that labs double check Shariant when a variant in Alissa is highlighted as existing in Shariant to view the latest records.

5.16.6 VCF

VCF (Variant Class Format 4.1) is a common format used for bioinfomatics.

- The VCF will have one row per variant, but using additional info fields have repeatable values. So if a single variant has three classifications, the INFO fields lab and chgvs will each have three comma separated values. The first lab and first INFO chgvs for that row relate to the same record. Likewise the second of each of those fields refer to the next record, etc.
- The VCF has the least amount of data of all the formats, but provides a link to the relevant Shariant page to see all data.