

VASA USER MANUAL

Revision Sheet

Release No.	Date	Revision Description
Rev. 0	26/06/2017	Document creation (draft)
Rev. 01	17/08/2017	Updates to editing records and current worklist sections
Rev. 02	17/10/2017	Include multiple worklist sections
Rev. 03	24/11/2017	Changes to API section
Rev. 04	12/06/2018	Changes to API section
Rev 2.0.0	01/06/2019	Major API changes. Resolved multiple assembly issues Changed revision number to match the VASA version number

Table of Contents

1.0	Introduction.....	3
1.1	System Purpose.....	3
1.2	Points of Contact.....	3
1.3	Compatibility	3
2.0	Getting Started.....	4
2.1	Logging On	4
2.2	Navigating the system.....	4
3.0	Importing a Variant List	6
3.1	Import VCF.....	6
3.2	Import a Sophia Genetics CSV / Mol Path TSV	7
3.3	Import a Single Variant	7
3.4	Import a Training Set	8
3.5	The Worklists	8
4.0	Scoring Variants	9
4.1	Current Worklist	9
4.2	Scoring a Variant	10
5.0	Viewing the scores database.....	16
5.1	All Variants View	16
5.2	Variant Details.....	16
5.3	All Scores View	17
5.4	Score Details.....	18
6.0	API Reference	19
6.1	Importing Single Variants	19
6.2	Token authentication.....	19
6.3	Submitting a worklist	20
6.4	Querying the database.....	21
7.0	Acronyms	23

1.0 INTRODUCTION

1.1 System Purpose

UK adoption of the American College of Medical Genetics and Genomics (ACMG) guidelines for genetic variant pathogenicity scoring initiated us to develop a tool to assist clinical scientists during this routine process.

VASA (Variant Scoring Assistant) is a web-based tool that guides a user through the process of classifying variants against the ACMG categories; recording evidence that supports decisions and calculating the pathogenicity score. The guidelines can be overridden, allowing scientists to exercise judgment where necessary.

VASA retains the calculations, variant annotations and associated evidence in a repository for reference in future submissions. It does not however, store patient specific data, allowing sharing of evidence between users.

Variants may be scored multiple times and all scores and associated evidence are available to all users. Training VCF datasets can be uploaded allowing trainers to review how trainees scored variants, viewing the evidence recorded and which ACMG categories caused any variation.

If you find any errors or have any feedback please contact philip.davidson2@nhs.net

1.2 Points of Contact

In the event of problems or questions please contact Philip Davidson on philip.davidson2@nhs.net.

1.3 Compatibility

VASA has been developed using Django as the web framework (version 1.11), and Twitter Bootstrap as the HTML & CSS framework (version 3), and as such should be compatible with most recent browsers.

At the time of writing v2.0.0, VASA has been testing with Google Chrome v.74, Mozilla Firefox v67 and Microsoft Edge 40.

2.0 GETTING STARTED

VASA is available at: <https://vasa.rosalind.kcl.ac.uk>

2.1 Logging On

Log in credentials can be obtained by emailing philip.davidson2@nhs.net. Please supply the name, email address and organisation of the user/s that require accounts.

2.2 Navigating the system

All functions are available from the navigation bar at the top every page (Figure 1).

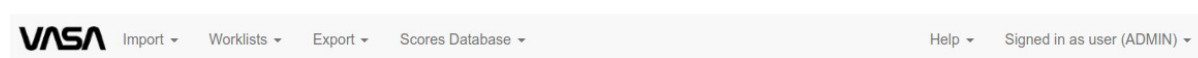


Figure 1. The navigation bar

Each function is covered in detail in sections 3 to 6, however a brief overview is given below.

2.2.1 Import

This link gives a drop-down list to several methods for importing variants to the system to create a 'worklist' of variants to be scored. The options are:

- Import VCF: upload a locally stored VCF file to create a worklist
- Import Single Variant: Manually enter a variant by genomic co-ordinates
- Import CSV or TSV: Import the proprietary CSV format of variants produced by the Sophia Genetics DDM[®] application or the KCH Molecular Pathology tab delimited variant lists
- Import Training Set: Create a worklist using one of the pre-loaded list of variants

If you are new to the system, choosing the 'example worklist' from the 'Import Training Set' will be the simplest way to get started and become familiar with the system.

Once a training set has been loaded, you will be redirected to the current worklist view described below.

2.2.2 Worklists

This displays either the current worklist or a list of previously imported worklists of variants being scored. The table displays the progress of scoring for each variant and links to initiate scoring.

2.2.3 Export

This page displays all the variants that have been scored in the current session by the current user. This list of variants can then be exported as a CSV file.

2.2.4 Scores Database

This links to existing scores contained within the VASA database.

2.2.5 Help

This contains links to the application details, the user manual and contact details for technical support.

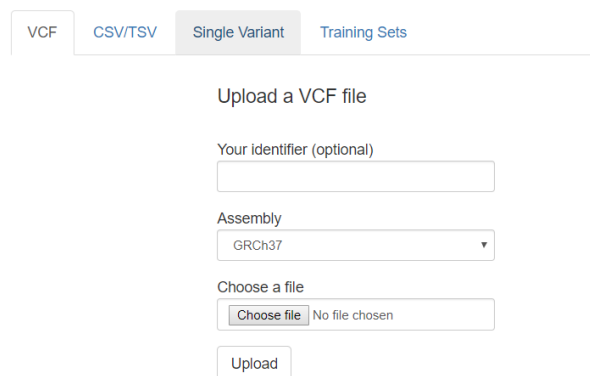
2.2.6 Login/Signed in as....

This is the link for logging in, or if you are already logged in, the link for logging out or changing your password.

3.0 IMPORTING A VARIANT LIST

A variant, or variant list, can be imported using several methods. This section describes how to manually upload variants to create a worklist. Uploading variants via the API is covered in section 6.

With the exception of the 'training set' import, you will be presented with a form with



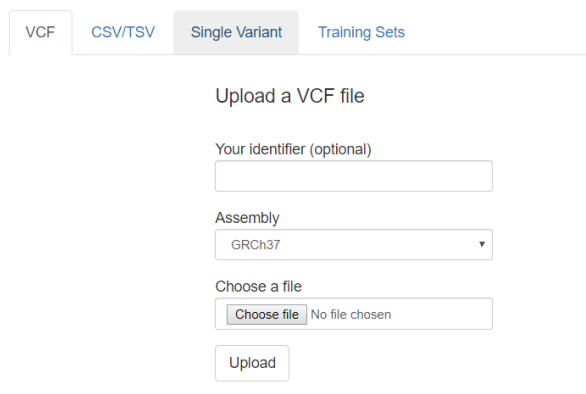
The screenshot shows a web interface with four tabs: 'VCF', 'CSV/TSV', 'Single Variant', and 'Training Sets'. The 'Single Variant' tab is selected. Below the tabs, the text 'Upload a VCF file' is displayed. There are three main input fields: 'Your identifier (optional)' (a text box), 'Assembly' (a dropdown menu with 'GRCh37' selected), and 'Choose a file' (a button labeled 'Choose file' next to the text 'No file chosen'). At the bottom is an 'Upload' button.

three fields as shown in

Figure 2. The common fields are described below.

Your identifier (optional): This field will be stored with the final variant score and it is intended to provide a link between the user's sample and the final score (i.e. a laboratory identifier) when searching the VASA database. It should not be used for patient identifiable information.

If this option is left blank, VASA will use the file name as the identifier if a VCF or CSV is uploaded, or simply the genomic coordinates in the case of an individual variant.



This is a duplicate of the form shown in Figure 2, displaying the 'Single Variant' import interface with tabs, input fields, and the 'Upload' button.

Figure 2. An import dialogue box

Assembly: This defaults to GRCh37 and GRCh38 are available

Choose a file: Enter the path of the local copy of the VCF file you wish to import.

Once a variant or list of variants have been successfully imported, you will be redirected to the worklist page (see section 4).

3.1 Import VCF

VASA uses PyVCF v.0.6.8 (<https://pyvcf.readthedocs.io>) to parse the VCF file and so supports VCF versions 4.0 & 4.1.

VCF Annotation

VASA has been designed to accept annotated VCF files using VEP annotation (http://grch37.ensembl.org/Homo_sapiens/Tools/VEP), and tested up Ensembl version 90 (including the new gnomAD annotation). Whilst this is optional, the advantage of using annotated VCF files is that the data will be available during the scoring process to users and will be stored with the final scores in the database. This can assist the user in the scoring process and provide an audit trail of the evidence that available at the time of scoring, without the need to manual input data. There is no defined set of expected VEP annotation fields; VASA will display and store whatever annotation fields are provided.

** If more than one annotation set is provided for a variant, an entry for each will be displayed in the worklist

3.2 Import a Sophia Genetics CSV / Mol Path TSV

Like the VCF file, VASA can be used to import proprietary text files of variants produced by the Sophia Genetics Sophia DDM application or the Molecular Pathology lab NGS pipeline. The CSV is created from the 'SNV/Indels' table of Sophia DDM (see the Sophia DDM user manual for more information). As with the VCF files, VASA will parse the associated variant annotation and store this with the variant.

3.3 Import a Single Variant

You can add a single variant for scoring for *ad hoc* tasks. The form requires the variant's genomic co-ordinates in the format:

[chromosome]:[position]:[ref_allele]:[alt_allele] e.g: 1:123456:A:T

VASA will make a request to Ensembl (<http://rest.ensembl.org/>) to confirm the validity of the co-ordinates and the reference allele. Only the validity of SNPs will be checked – due to the ambiguity of indel co-ordinates, these will be passed through unchecked.

There is also the option to add the transcript ID. Currently there is no validation of the field, but it will be stored with the variant score.

VASA does not annotate a variant import using this method as this is the responsibility of the user. If annotation is required, it is suggested that VEP is used to create an annotated VCF file and the instructions in section 3.3 can be used.

Possible Import errors:	
Please check the format of the genomic co-ordinates	The variant has not in the correct format described above
Unable to parse HGVS notation	The variant is correctly formatted but is outside of the expected range (i.e. an invalid chromosome number or co-ordinate)
Reference allele extracted from xxx does not match reference allele given by HGVS notation	The co-ordinates are valid but the reference allele is not the one expected. If this occurs, in the first instance check the use of the correct assembly

3.4 Import a Training Set

VASA has several built-in sets of variants, both for use as training tool for the application itself, or for training on scoring variants with the ACMG guidelines. If you are new to the system, the 'example worklist' will be the good place to get started and become familiar with the system.

Most training sets will come with information at the top of the worklist page as additional instructions, such as clinical background in the case of hypothetical training cases.

Training set data will not be submitted to the main database, therefore not impacting on clinical datasets.

3.5 The Worklists

VASA will store worklists to allow users to swap between them and working on more than dataset at a time (Figure 3). Any worklist that is deleted will not impact on variants already submitted to the database, but you will lose any partially scored variant data.

Your saved sessions

	Session ID	Date Created	Date Modified	Your identifier
load purge	17	17 Oct 2017, 12:18 p.m.	17 Oct 2017, 12:18 p.m.	vep_test
load purge	16	17 Oct 2017, 12:18 p.m.	17 Oct 2017, 12:18 p.m.	vep_test
load purge	15	17 Oct 2017, 12:18 p.m.	17 Oct 2017, 12:18 p.m.	vep_test
load purge	14	17 Oct 2017, 12:18 p.m.	17 Oct 2017, 12:18 p.m.	vep_test
load purge	13	17 Oct 2017, 12:18 p.m.	17 Oct 2017, 12:18 p.m.	vep_test
load purge	11	17 Oct 2017, 9:47 a.m.	17 Oct 2017, 11:35 a.m.	vep_test
load purge	10	17 Oct 2017, 9:47 a.m.	17 Oct 2017, 9:47 a.m.	vep_test
load purge	9	17 Oct 2017, 9:47 a.m.	17 Oct 2017, 9:47 a.m.	vep_test
load purge	4	17 Oct 2017, 9:39 a.m.	17 Oct 2017, 9:39 a.m.	EQA Scheme 2017

Figure 3. The user's worklists stored in VASA

4.0 SCORING VARIANTS

4.1 Current Worklist

Once a list has been imported, you will be redirected to the worklist view. This lists all the imported variants, any associated annotation and links to score the variant and show if this variant has been seen before by VASA. This is shown in

Figure 4.

The screenshot shows a table with four columns: 'To Do', 'Org score', 'Global score', and 'Alt. transcripts?'. The table contains six rows of variant data. Annotations with arrows point to specific cells or rows:

- 'This has been scored previously but your org score needs revisiting' points to the 'Accept score' button in the first row.
- 'This has been scored previously but your org score has not been verified' points to the 'Accept score' button in the second row.
- 'This has been seen before but **not** score by your organisation' points to the 'Score' button in the third row.
- 'Your org score is valid but there is discordance between organisational' points to the '3.0 / 1' global score in the second row.
- 'Completely novel variant to VASA' points to the '2.0 / 2' global score in the fourth row.
- 'Has been seen before but scored against a different transcript' points to the 'view' button in the sixth row.

To Do	Org score	Global score	Alt. transcripts?
Accept score	3	3.0 / 2	-
Accept score	3	3.0 / 1	-
Score	-	3.0 / 1	-
Accept score	3	2.0 / 2	-
Score	-	-	-
Score	-	-	view

Figure 4. The current worklist view. The variant data and annotation has been cropped, and only the first four columns have been shown here.

The To Do column

The worklist will keep track of the variants you score. The 'Score' button allows you to start work on a variant and 'In progress' shows which variants have already been started. 'Done' indicates a variant has been scored and submitted to the database in this session.

If a variant has been previously scored by your organisation, there will be the option to 'accept score'. Choosing this option will not change anything in the database, but will simply allow you to track your progress within the worklist.

VASA will allow you to work on multiple variants and retain progress on each. Progress will be kept even if you log out and/or close the browser.

The ORG Score column

This will show if a variant has been scored previously by your organisation, along with the value (1 = benign to 5 = pathogenic).

The button is also colour-coded:

- green means score has been second checked,
- red means no second check
- amber means needs review (the second check was > 365 days ago).

Clicking the button will take you to the scoring details and the potential to edit if required.

Global Score column

This has been scored previously by any organisation. The first number is an average of the scores (1 = benign to 5 = pathogenic). The second number is how many times it has been scored.

In addition the button is colour coded:

- green means all the scores in VASA match
- amber means there is discordance.

Alt. transcripts? column

If the same variant has been scored (against a different transcript), this will link to the record/s.

4.2 Scoring a Variant

From the worklist view, click on the white 'Score' button to start scoring your chosen variant (or 'In Progress' to continue with a variant has been partially scored).

4.2.1 Category Dashboard

This will take you to the 'category dashboard' showing the 28 ACMG categories and other summary information, as shown in Figure 5.

The screenshot shows a web interface for variant scoring. On the left, there are three tabs: 'Categories', 'Variant Annotation', and 'Existing literature'. Below the tabs, evidence is categorized into three groups: 'Population databases', 'Database evidence', and 'Functional Studies'. Each category contains a list of items, each with a 'Pending' status (orange button) and an 'Annotation' button (grey button). For example, under 'Population databases', there are five items: BA1 (Allele frequency is >5% in controls), BS1 (Allele frequency is greater than expected for disorder), PM2 (Absent from controls), PS4 (The prevalence of the variant in affected individuals is greater than controls), and BS2 (Observed in a healthy adult for a disorder with full penetrance). Under 'Database evidence', there are four items: PS1 (Same amino acid change as a previously established pathogenic variant), PM5 (Novel missense at a codon where a pathogenic missense change has been seen), PP5 (Reputable source reports variant as pathogenic, but the evidence is not available), and BP6 (Reputable source recently reports variant as benign, but the evidence is not available). Under 'Functional Studies', there are two items: PS3 (Well-established in vitro or in vivo functional studies supportive of a damaging effect) and BS3 (Well-established in vitro or in vivo functional studies show no damaging effect). On the right, a rounded rectangular panel titled 'Submit (option to review)' contains a 'Provisional score' section showing 'Uncertain significance'. Below this are fields for 'Phenotype' and 'Alias', a dropdown for '*Override final category?' set to 'No, leave as calculated', and a 'Notes' text area. At the bottom of the panel are 'Save Changes' and 'Cancel' buttons.

Category	Status	Evidence	Action
Population databases	Pending	BA1: Allele frequency is >5% in controls.	Annotation
	Pending	BS1: Allele frequency is greater than expected for disorder.	Annotation
	Pending	PM2: Absent from controls (or at extremely low frequency if recessive).	Annotation
	Pending	PS4: The prevalence of the variant in affected individuals is greater than controls.	
	Pending	BS2: Observed in a healthy adult for a disorder with full penetrance.	
Database evidence	Pending	PS1: Same amino acid change as a previously established pathogenic variant.	
	Pending	PM5: Novel missense at a codon where a pathogenic missense change has been seen.	
	Pending	PP5: Reputable source reports variant as pathogenic, but the evidence is not available.	Annotation
	Pending	BP6: Reputable source recently reports variant as benign, but the evidence is not available.	Annotation
Functional Studies	Pending	PS3: Well-established in vitro or in vivo functional studies supportive of a damaging effect.	
	Pending	BS3: Well-established in vitro or in vivo functional studies show no damaging effect.	

In Silico evidence

Figure 5. The category dashboard

All the categories will start as 'Pending' (coloured orange) and you can work through answering as many or few categories as you consider relevant.

The categories are grouped by type of evidence (population database, functional studies, family data, etc.).

The overview panel to the right of the screen shows the current score based on the categories that have been completed. As per the ACMG guidelines, a variant defaults to 'Uncertain Significance' in the absence of evidence of pathogenicity or benignity.

The Phenotype field

This field is compulsory, and is a free-text field that will populate with HPO terms as suggestions. Multiple HPO terms can be entered.

Alias

This is an optional field for noting any alternative, historic names for the variant

The screenshot shows the 'Overview panel' on the right side of the interface. It contains a 'Provisional score' section with 'Uncertain significance'. Below this are fields for 'Phenotype' and 'Alias', a dropdown for '*Override final category?' set to 'No, leave as calculated', and a 'Notes' text area. At the bottom of the panel are 'Save Changes' and 'Cancel' buttons, and a 'Review and submit' button.

Figure 6. The Overview panel.

Override final category?

It is possible for the scorer to override the final variant score if required. If this is done, the score will be displayed with an asterisk as shown below:

Notes can also be added that will be stored with the variant scoring.

4.2.2 Category Questions

Clicking on the any category 'Pending' button will take you to the page for storing information for each category. This is shown below in for the first category, BA1. The top panel gives the full text for the category from the ACMG guidelines, with two panels below that. The lower left panel is for recording information relating to the category and the lower right panel will display variant annotation (if available) that is pertinent to the category in question.

Variant Annotation	
AF	0.2280
AFR_AF	0.0658
AMR_AF	0.4265
EAS_AF	0.2004
EUR_AF	0.3141
SAS_AF	0.2464
AA_AF	0.1071
EA_AF	0.3221
ExAC_AF	0.284
ExAC_Adj_AF	0.2843
ExAC_AFR_AF	0.104

Figure 7. The category question page

Have the criteria been met?

This gives three options; 'yes', 'no' and 'don't know'. Checking 'yes' will mark this criterion as met, with regards to calculating the variant's pathogenicity. The other two options will **not** change the scoring of the variant, but will allow the user to (optionally) record that the category has been considered and either not met or there is insufficient evidence. This may be useful in more complex cases where the scorer wished to record that they have considered the criterion but believe it has not been met.

Weighting

This dropdown list allows the option to change the default weighting for the category, as there may be some occasions where this is required. The new weighting for the category will be applied when the pathogenicity score is calculated.

Evidence

This is an optional free-text field for recording comments relevant to this category.

Once the category information has been recorded, clicking submit will return you to the dashboard and the category will be marked as changed, as shown below. You can return to any category to amend or update information until the final submission.

The screenshot shows a web interface with two tabs: 'Categories' and 'Variant Annotation'. The 'Categories' tab is active. Below the tabs, there is a section titled 'Population databases'. It contains two rows. The first row has a blue button labeled 'True' and the text 'BA1: Allele frequency is >5% in controls.' The second row has an orange button labeled 'Pending' and the text 'BS1: Allele frequency is greater than expected for disorder.'

Figure 8. The display will change once a question has been completed.

4.2.3 Variant Annotation

The second tab on the dashboard allows you to view all the variant annotation that was imported with the variant. This may be just the core variant details (chromosome, genomic position, etc.) or more depending on the options at import.

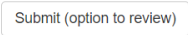
Scoring for COL4A5 (NM_000495.4): c.4510+5G>C

The screenshot shows the 'Variant Annotation' tab. It displays a table with the following data:

Identifier	Case 4
CHROM	X
POS	107930929
REF	G
ALT	C
GENOMIC_POS	X:g.107930929G>C
display_name	COL4A5 (NM_000495.4): c.4510+5G>C
Consequence	splice_region_variant&intron_variant
VCF ID	X:g.107930929G>C
IMPACT	LOW
SYMBOL	COL4A5
Gene	1287
Feature_type	Transcript
Feature	NM_000495.4
BIOTYPE	protein_coding
EXON	
INTRON	47/50

Figure 9. Variant Annotation

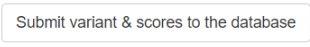
4.2.4 Reviewing and Submitting the Score

Once you are satisfied the scoring is complete, click  the button.

This will take you to a page with the option to review the data before submission.
This will be summarised as below:

Review data for FGFR3 (NM_000142.4): c.1138G>A p.Gly380Arg

Summary:

Score:	Pathogenic	
Patient Phenotype:	Frontal bossing, Lumbar hyperlordosis, Limb undergrowth, Macrocephaly	
Identifier:	achondroplasia.vcf	
Notes:		

Categories Completed:

Category	Criteria met?	Weighting	User Notes
PM2	True	pm	
PS4	True	ps	
PP3	True	pp	
BA1	False	ba	
PS3	True	ps	Well established causative variant for Achondroplasia
PM3	True	pm	

Figure 10. The review page.

Nothing will be committed to the database until the 'Submit variant' button is clicked. VASA will check if the variant is already present in the database, and if so, it will associate the new score with that variant, else if the variant is new to the database it will submit both the variant and score data.

Once submitted, you will be returned to the 'current worklist' page.

4.2.5 Amending an existing record

When viewing an existing record during the scoring process, if the score was submitted by your organisation, there will be an 'edit record' button at the top of the page. This will open the record in the scoring dashboard again and allow you to amend the record and resubmit.

Scoring details for 11:g.35314059G>C

[Edit Record](#)[Confirm scoring \(2nd check\)](#)**Summary:**

Score:	Uncertain significance
Patient phenotype:	vhbhb
Alias:	
Submitted by	KCH on 17 Aug 2017, 1:41 p.m.
Last scored:	Phil Davidson, 17 Aug 2017, 1:41 p.m.
Checked:	-
Identifier:	11:g.35314059G>C

Notes:

Figure 11. Editing a record

If the record has not been second checked (and you were not the original scorer), there will be a second button for you to confirm the data and you will be marked down as the second checker for the record.

- If you choose to edit a record that you did not originally submit, the record will be amended to record you as the new scorer of the record.
- If the record had been second checked, this status will be cleared and will now need a second check from another user in your organisation. Thus, any changes to a record will now need a second person to confirm those changes.

5.0 VIEWING THE SCORES DATABASE

There are two basic views for querying the data: 'all variants' and 'all scores'. These are accesses from the 'scores database' dropdown list on the navigation bar. Although activity such as scoring a variant is logged against a user in the database, VASA will only display the user's organisation (not the named individual) to database viewers from outside that organisation.

5.1 All Variants View

A screenshot of the variants table is shown below. It gives the genomic coordinates for each variant and summary information such as how many scores are present in the database, the average score and when the last time it was score was submitted.

All Variants

Search variants: <input type="text"/> <input type="button" value="Search"/>					
ID ▲	Genomic coords ↕	Assembly ↕	Average score ↕	Most recent ↕	Total times scored ↕
464735	2:8923281:C:T	grch37	1.0	24/11/2015	1
464736	1:8925474:G:A	grch37	1.0	24/11/2015	1
464737	1:11847436:T:C	grch37	1.0	12/07/2016	1
464738	1:11854457:G:A	grch37	1.0	22/08/2016	1
464739	1:11856311:G:A	grch37	2.0	24/11/2015	1
464740	1:12387766:C:T	grch37	1.0	24/11/2015	1
464741	1:29442291:C:T	grch37	1.0	02/03/2016	2
464743	1:43394666:G:A	grch37	1.0	24/11/2015	1
464744	1:43394887:G:A	grch37	1.0	24/11/2015	1
464745	1:45974849:A:G	grch37	1.0	24/11/2015	1
464746	1:47685569:G:A	grch37	1.0	24/11/2015	1
464747	1:47691350:C:T	grch37	2.0	24/11/2015	1
464748	1:47696581:C:A	grch37	1.0	13/11/2015	1

Figure 12. All Variants view

There is a link to view more details for the variant. This is described in section 5.2

5.2 Variant Details

Selecting one of the variants from the 'All variants' table will give more detailed data on that variant, including when it was first submitted to the database and by whom. There are three tabs under each variant

5.2.1 Scoring statistics

This shows how many times the variant has been scored, and the average and most frequent score.

Detail for variant 15:43500977:T:C

Summary:	
Assembly	grch37
Date submitted	02/05/2018
First submitted by	KCH (admin)
Times scored:	1
Average score:	1.0
Most recent score:	21/04/2016
Most frequent:	Benign

Individual Scores
Category breakdown

Score	Transcript	Date	Submitted by	Against phenotype	Identifier	ACMG?	Checked
Benign	NM_000119.2	21/04/2016	KCH	None			⚠ Expired

Showing 1 to 1 of 1 entries

Figure 13. Variant Details with the list of previous scores per variant

5.2.2 Category Breakdown

This shows a summary of which categories have been marked as met across all the scoring. This will give a useful insight into the origins of variation in the final scoring of variants:

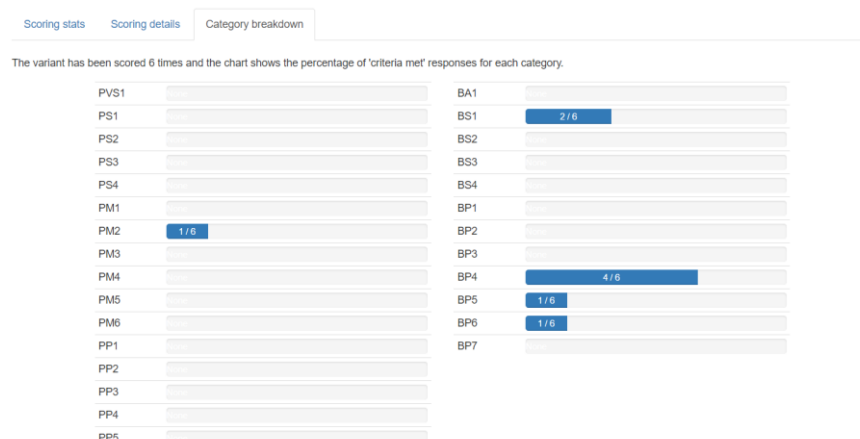


Figure 14. Category summary breakdown

5.3 All Scores View

Like the 'All Variants' view, this view shows all scores submitted to the database. This table can be useful in seeing the most recent activity within the database. Links to the related variants are available (detailed in section 5.2) and to the score details, (described in the section 5.4). The final column shows if the score has been second checked, and if that check has expired. The expiry time is currently set to 1 year as a global setting.

Show 15 entries

Search:

ID	Score	Variant	Transcript	Phenotype	Identifier	Date Added	last scored by	ACMG rules?	Checked
187	Likely pathogenic	1:g.158651385G>GCAA	NM_003126.2	Elliptocytosis	Concordance analysis data set - TSV	11/09/2017	KCH	✓	✓ Yes
198	Benign	1:g.8931463T>C	ENST00000234590	Myotonia of the jaw	1:g.8931463T>C	15/11/2017	KCH	✓	✓ Yes
1	Benign	13:g.110859069G>C	None	Cataract	initial	14/05/2017	STG	✓	⚠ No
3	Pathogenic	4:g.1806119G>A	None	Achondroplasia	initial	15/05/2017	KCH	✓	⚠ No
4	Benign	12:g.9013755C>T	None	Rasopathy Panel	initial	15/05/2017	KCH	✓	⚠ No

Figure 15. All Scores View

5.3.1 Scores needing a second check

This is a view of the data similar to the 'all scores view' but restricted to only scores from your organisation that require a second check. This view is here to allow a user to go through and systematically second check those that need it, independent of the whole import and worklist workflow.

5.4 Score Details

This page gives the details for each submitted score, including the audit information such as the scorer and date and time plus all the submitted categories answers and user submitted evidence. If it was provided, it will also store the variant annotation available at the time. This should provide useful as some annotation may change over time, it will be possible to see a snapshot of the evidence provided at the time of scoring.

Summary:

Score:	Likely pathogenic
Patient phenotype:	Affected
Alias:	None
Submitted by	STG on 1 Jun 2017, 12:41 p.m.
Last scored:	20 Jul 2017, 10:19 a.m.
Checked:	⚠ No
Identifier:	ACGS Cases 4-9
Notes:	

ACMG Categories Completed:

Category	Criteria met?	Weighting	User Notes
BA1	False	-	
PM1	True	pm	
PP3	True	ps (Changed from default)	

Figure 16. Score details

6.0 API REFERENCE

A basic API is available to provide a simple, automated method integrating into your existing workflows. This can be broken down into two functions: importing variants for scoring, and querying the database.

6.1 Importing Single Variants

A simple GET method for importing single variants is available.

Using the URL:

```
https://vasa.rosalind.kcl.ac.uk/imports/get_variant/[assembly]/[genomic_pos]
```

e.g:

```
https://vasa.rosalind.kcl.ac.uk/imports/get\_variant/grch37/1:123456:A:T
```

This will redirect the user to the 'current worklist' view with the single variant ready for scoring. Note, this method can be used if you wish to generate a URL within another application that takes a user directly to a single variant worklist and requires the user to be logged in to VASA. If the user is not logged in, they will be redirected to the log in page first, then to the worklist after successfully entering their log in credentials.

6.2 Token authentication

For the following sections, 6.3 & 6.4, simple user token authorisation has been implemented. The request will need to contain the authorisation in the header:

```
Authorization: Token <user's token>
```

A user's token can be found in two ways. Either from the GUI or programmatically. From the website, click on the 'signed in as...' in the menu bar, then 'view user profile'. To retrieve the API token programmatically.

Using a command line http client such as httpie:

```
https POST vasa.rosalind.kcl.ac.uk/api-token-auth/ username='user' password='password'
```

or with Python 'requests module':

```
import requests

url = 'https://vasa.rosalind.kcl.ac.uk/api-token-auth'
response = requests.post(url, data={
    'username': '[user]',
    'password': '[password]'})

print(response.json())
```

VASA will return the user's token in a JSON response:

```
{
  "token": "<user token>"
}
```

6.3 Submitting a worklist

You can submit a complete worklist to VASA, including any annotation. Below is a simple python script demonstrating the process:

```
import requests, json

url = 'https://vasa.rosalind.kcl.ac.uk/rest/post_worklist/'
auth_token = 'Token <your token>'
headers = {'Authorization': auth_token, 'content-type': 'application/json'}

worklist = {
  "run_name": "my_worklist",
  "assembly": "grch38",
  "annotation_type": "vep",
  "variants": [
    {
      "chromosome": "1",
      "position": "196642221",
      "ref_allele": "T",
      "alt_allele": "G",
      "transcript": "NM_000186.3",
    },
    {
      "chromosome": "11",
      "position": "35314059",
      "ref_allele": "G",
      "alt_allele": "C",
      "transcript": "NM_004171.3"
    }
  ]
}

response = requests.post(url, data=json.dumps(worklist), headers=headers)
print(response.json())
```

VASA returns the id of the worklist in a simple JSON format, thus allowing the creation of a URL to redirect the user within the requesting application:

```
{'url': 'http://[hostname]/imports/worklist/xx'}
```

6.3.1 Worklist annotation types

The example above only includes the core information but as many annotation fields can be included as desired. The 'annotation_type' is used by VASA to determine which ACMG category each piece of evidence is relevant to, and when it is displayed during the scoring process.

The accepted values are: 'vep', 'mol_path_csv' & 'custom'.

'vep' and 'mol_path_csv' are annotation sets that are known to VASA. If you wish to submit your own, use 'custom' and supply an additional field in the data as shown below, mapping the annotation field name to the ACMG category:

```

worklist2 = {
  "run_name": "worklist2",
  "assembly": "grch37",
  "annotation_type": "custom",
  "cat_map": {
    "ps1": ["Exac", "consequence"],
    "PS2": ["Feature", "consequence"],
    "BS2": ["Exac", "consequence"]},
  "variants": [
    {
      "chromosome": "1",
      "position": "966424291",
      "ref_allele": "T",
      "alt_allele": "G",
      "transcript": "NM_000186.3",
      "consequence": "missense_variant"},
    {
      "chromosome": "11",
      "position": "13539144059",
      "ref_allele": "G",
      "alt_allele": "C",
      "transcript": "NM_004171.3",
      "consequence": "missense_variant"}
  ]
}

```

6.4 Querying the database

The Django REST Framework has been used to create the API for querying data. Please refer to section 6.2 for obtaining your API token.

Individual variant records can be retrieved with:

[https://vasa.rosalind.kcl.ac.uk/rest/variant_detail/\[assembly\]/\[genomic_pos\]/?format=json](https://vasa.rosalind.kcl.ac.uk/rest/variant_detail/[assembly]/[genomic_pos]/?format=json)

Individual scores can be retrieved with:

[https://vasa.rosalind.kcl.ac.uk/rest/score_detail/\[score_index\]/?format=json](https://vasa.rosalind.kcl.ac.uk/rest/score_detail/[score_index]/?format=json)

The format for records can be seen on the following page.

```

"id" : int,
"chromosome" : string,
"position" : string,
"ref_allele" : string,
"alt_allele" : string,
"genomic_pos" : string,
"assembly" : string,
"date_created" : string,
"submitted_by" : string,
"organisation" : string,
"request_date" : string,
"requested_by" : string,
"times_scored" : int,
"score_mismatch" : Boolean,
"scores": [
  {
    "id" : int,
    "final_score" : string,
    "auto_score" : string,
    "phenotype" : [List],
    "user_notes" : string,
    "scored_by" : string,
    "date_scored" : string,
    "checked_by" : string,
    "date_checked" : string,
    "organisation" : string,
    "date_submitted" : string,
    "request_date" : string,
    "requested_by" : string,
    "transcript" : string,
    "variant_assembly" : string,
    "variant_string" : string,
    "scored_by_acmg" : boolean,
    "scoring_data" : {
      string : {
        "user_text" : string,
        "weighting" : string,
        "category" : string,
        "criteria_met" : True|False|N/K
        "initial_weighting" : string
      }
    },
    "variant_annotation" : {
      Key: Value,
    },
  },
],
]

```

Variant index number in VASA
Following columns refer to variant submission in db

in '1:g.123456A>T' format

Date that variant was first submitted to VASA, in ISO8601 format
User that first submitted variant
Organisation that first submitted variant
Date that request is made, now(), in ISO8601 format
Username of the requester
How many scores exist for this variant
If there is >1 score for this variant, is there a mismatch?
'scores' is a list: multiple scores per variant

Score index number in VASA
Final score (the 'auto_score' may have been overridden)
Automatic scoring based on ACMG criteria

Username of last scorer ('not_available' if it's from a different organisation)
Date that variant was scored, in ISO8601 format
Username of checker. Null if not checked, 'not_available' if it's from a different org)
Date that variant was checked, in ISO8601 format. Null if not checked
Organisation that the score was submitted by
Date score was first submitted
Date that request is made, now(), in ISO8601 format
Username of the requester
Transcript that the variant was scored against

If false the following scoring_data field is ignored
scoring_data holds ACMG category data. A child object for each category
The category name, e.g. 'PVS1', 'BA1'

e.g. pvs, ps, pp, ba, (this may have been changed from default)

e.g. pvs, ps, pp, ba, (the default, not necessarily the value used)

Variant annotation: any number of key/value pairs. This is per score, not variant

7.0 ACRONYMS

CSV	Comma Separated Values
VEP	Ensembl's Variant Effect Predictor
VCF	Variant Call Format
ACMG	American College of Medical Genetics and Genomics
AMP	Association for Molecular Pathology
HGVS	Human Genome Variation Society